

Alessio Montagner

IL RESPIRO DEL PENSIERO

Introduzione ad una filosofia analitica in otto ragionamenti



Presentazione

Nella primavera del 2017 ero all'apice della mia esistenza, credevo di aver messo distanza tra me e tutte le preoccupazioni e minacce. Ero studente di lettere moderne a Ca' Foscari (come in teoria sono ancora, e Dio solo sa se potrò mai finire) ma mi interessavo anche di altro, stavo studiando matematica e chimica per conto mio, stavo riflettendo su cosa fare del mio futuro e credevo che fosse segnato per me il diventare professore (anche se il mio sogno era fare l'architetto del paesaggio). Ero felice.

A volte, la sera, buttandomi a letto, vedevo il buio avanzare, mi sentivo scomparire, venivo colto dal *terrore arsamasiano* come nella Confessione di Tolstoj (2009). Poi, un giorno come gli altri, non sono più riuscito a liberarmene.

Non che mi preoccupasse il mio sparire, mi preoccupava lo sparire degli altri; anzi, non mi preoccupava neanche il loro sparire, mi preoccupava il loro spegnersi, l'ingiustizia che questa energia vitale potente si sarebbe esaurita.

Non ero venuto a conoscenza di nulla di nuovo, erano solo cambiate le mie emozioni. Di colpo mi sembrava assolutamente folle il fatto che la gente riuscisse a lavorare, mangiare, vivere come se niente fosse, come se non si rendesse conto del problema enorme della morte. Sapevo, intellettualmente, che la morte non era imminente per me e i miei cari; ma, emotivamente, la sua inevitabilità già me ne faceva sentire il rantolo, sentivo come se quel momento, il momento della fine, fosse lì a pochi secondi. Ho passato mesi a piangere disteso sul divano senza che nessuno se ne accorgesse.

Con Seneca (2015) voglio dire: «è grazie alla filosofia se mi sono risollevato, se sono guarito: alla filosofia sono debitore della vita, ma questo è il debito più piccolo che ho con lei»; e non perché mi abbia dato chissà quale speranza o risposta, ma perché la sua concettosità, l'insopportabile astrattezza, l'incomprensibilità di certe diatribe e l'assurdità di certe domande hanno occupato la mia mente scacciando gli altri pensieri diretti verso il nulla.

Per questo ho tanto desiderato scrivere questo libro.

Cosa ho scritto?

Non ho voluto scrivere un testo accademico, per filosofi professionisti. Non avrei potuto, io non sono un filosofo, non sono niente. Quella del *filosofo professionista* è un'invenzione recente, un tempo la filosofia era tutto ed era praticata da tutti, avvocati, medici, poeti. Wittgenstein, il più grande filosofo del Novecento, quando un alunno gli confessava che avrebbe voluto *fare il filosofo* lo rimproverava battendo i pugni, gli diceva piuttosto di fare il medico; lui stesso, nonostante fosse già professore in una delle maggiori università del mondo, voleva mollare tutto e iscriversi a medicina (o, in alternativa, fare l'operaio in qualche fabbrica sovietica). Anche nel secolo scorso abbiamo avuto contributi da filosofi non professionisti: Gustave Thibon, filosofo e teologo nominato al Nobel, era un contadino senza diploma; Dario Bernazza, del quale un gran filosofo come Ferraris dice che aveva capito Nietzsche meglio dei professori, era un imprenditore; Ayn Rand, scrittrice, ha creato il suo sistema di pensiero lavorando sempre al di fuori dell'accademia. Quella del filosofo è una professione che si auto-rifuta: se uno inizia a filosofare su cosa deve fare della sua vita conclude subito che non deve fare il filosofo.

Non ho voluto scrivere neanche una storia della filosofia. Sì, qui c'è una ottima fetta delle più importanti domande della filosofia analitica, e per ogni domanda le più importanti risposte proposte, e per ogni risposta i più importanti e stimolanti argomenti a favore e contro. Ma non è una esposizione completa, distaccata e neutrale: sarebbe stato futile provarci, qualsiasi libro ne sarebbe venuto fuori sarebbe stato surclassato dal primo manuale scritto da una manciata di specialisti in grado di dividersi il carico di lavoro.

E non ho voluto scrivere neppure una ricerca. Lo schema di questo libro è nato durante la quarantena del marzo 2020, è stato poi steso in pochi mesi quasi senza aprire un testo, senza spulciare riviste accademiche o richiedere testi alle biblioteche. Ho scritto invece ciò che ricordavo di aver letto, è stato tutto un lavoro di memoria, le citazioni sono state aggiunte in seguito. E questo mi dice qualcosa.

Mi piace la religione, mi piacciono tutte le religioni, mi piace sentirne l'odore, il sapore, vedere la devozione, la speranza, la santità. Ho una passione per gli ordini religiosi, per i loro vestiti e i loro carismi. Ho notato che gli ordini possono essere divisi in tre gruppi a seconda della loro attività distintiva:

- ci sono i teologi, come i domenicani, tutti concentrati sulla riflessione astratta;
- ci sono gli apostoli della carità, come i vincenziani, persone pratiche sempre con le mani protese verso i bisognosi;
- e ci sono i mistici, come i carmelitani, coloro che cercano di sperimentare Dio.

Mi colpiva che i tre gruppi rappresentassero i tre momenti dell'empirismo: la mistica è la percezione dell'oggetto, la teologia la rappresentazione e l'analisi mentale dell'oggetto, la carità l'azione muscolare risultante. Ma c'è dell'altro: i teologi, che sono filosofi, svolgono il loro lavoro con il pensiero; gli apostoli della carità, come artigiani, con le mani; ma i mistici, invece, lavorano di memoria, ricordano la loro esperienza e la trascrivono.

Quella che ho cercato di scrivere, allora, è un'opera di mistica: per questo è in prima persona; ho voluto scrivere di *avventure intellettuali*, di pensieri vissuti; una sorta di diario, volendo, che è l'unico genere di testo che posso scrivere.

Alla fine mi sono ritrovato con un'opera molto più ampia di quanto pensassi. Credevo avrei scritto un libretto agile, da tenere sul comodino o da leggere nei momenti vuoti in treno. L'infinità dell'universo non mi ispira meraviglia, mi ispira l'emozione di una catastrofe diluviana osservata da sotto la cappa di un camino, perché io sono fatto per le piccole cose, per il restare chiuso nella cabina della funivia in mezzo alla tormenta. E ho scritto anche molto più a lungo di quanto pensassi: pensavo di finire nel giro di un paio di mesi, e invece, dopo un anno, sono ancora qua che scrivo, e correggendo parola per parola mi sento soffocare ad ogni frase.

Perché ho scritto così tanto? Perché volevo capire cosa io stesso penso su questi temi, volevo capire la struttura della mia mente, lo stile del mio pensiero, e poterlo mostrare ad altri. E perché volevo che anche chi legge potesse capire qualcosa su di sé. Questo è un testo personale ma non è un testo diretto a filosofi, l'ho voluto scrivere per tutti: i ragionamenti sono esposti senza dare alcuna conoscenza per scontata, e come scrivere ha aiutato me a pensare, così desidero che chi legge possa pensare con me.

Individuo dunque tre interpretazioni. In primo luogo, questo libro è un'introduzione alla filosofia ma *ad una* filosofia: per esempio, quando parlo della verità, non espongo la teoria in modo neutrale,

ma difendo il realismo; cioè in ogni argomento presento la mia prospettiva personale, nella speranza di riuscire a dimostrare che la mia vita non è stata buttata anche se non ho i soldi per studiare. In secondo luogo, però, questo libro è anche un'introduzione generale: infatti, pur difendendo posizioni specifiche, comunque la loro discussione è abbastanza ampia da portarmi ad esporre tutte le più importanti argomentazioni per ogni campo della filosofia analitica, e dare così un affresco di tutta la disciplina. In terzo luogo, questo libro è infine un programma per future ricerche: tutte le cose che ho scritto devono essere approfondite, è necessario aggiornarsi sulle ultime pubblicazioni in modo tale da difendere le posizioni o ritrattarle.

Il libro si apre con una brevissima storia della filosofia che si concentra sul confronto tra la filosofia analitica e gli altri approcci contemporanei.

Seguono otto ragionamenti.

I primi due riguardano le domande dell'epistemologia, il settore che si occupa dei problemi della conoscenza.

I due successivi riguardano le domande dell'etica, il settore che si occupa del valore delle azioni e cerca di rispondere alla più grande delle domande: *come dobbiamo vivere*.

Le ultime quattro riguardano le domande della metafisica, il settore che si occupa dei problemi dell'essere, cioè si chiede qual è la natura delle cose, cosa c'è nel mondo.

Mi prendo tutta la responsabilità degli errori che si riscontreranno (e chi legge se lo ricordi, perché sarà rilevante).

Dedico questo libro ai miei avi e ai miei discendenti. Ai miei avi, soprattutto alle mie nonne Elena e Redenta, perché la filosofia mi ha donato molti rimorsi. E ai miei discendenti perché, se vorranno sapere da dove vengono, e se la mia famiglia conserverà una copia di ciò che ho scritto, potranno vedere un po' cosa pensava questo loro antenato.

Introduzione: perché la filosofia analitica non esiste

Un po' di storia...

Nella filosofia contemporanea si individuano di solito due tradizioni: quella analitica e quella continentale.

Quando uso questi termini ho ancora l'istinto di pensare subito a movimenti filosofici molto recenti, caratteristici del Novecento se non direttamente degli anni Cinquanta o Sessanta. In realtà è una distinzione che può essere fatta risalire a J. S. Mill (1840), il grande genio dell'etica dell'Ottocento. Ma proprio perché sono tradizioni, e non invenzioni, la loro origine è ancora più lontana...

Il primo passo: razionalisti ed empiristi

So che lo dirò molto spesso, che filosofia e matematica formano un tutt'uno: Talete, il primo filosofo; Pitagora, colui che per primo ha usato questa parola per descriversi; Platone, la fonte da cui sgorga tutta la filosofia occidentale, possono essere considerati tutti sia filosofi che matematici. Anche la filosofia moderna inizia, agli albori del diciassettesimo secolo, con un pensatore che giganteggia non solo tra i filosofi, ma anche tra i matematici di tutti i tempi: René Descartes, Cartesio.

Mentre i filosofi medievali, seguendo le orme di Aristotele, innalzavano il loro edificio filosofico mettendo alle fondamenta la metafisica (perché l'essere è la categoria più generale), Cartesio capisce che non può dire di conoscere alcunché se prima non determina cosa davvero sia la conoscenza e come la si possa ottenere: questa è la grande rivoluzione, l'epistemologia diventa il fondamento di tutto. La sua riflessione parte da un *dubbio iperbolico*, dice che pare non poterci essere alcuna certezza perché un demone potrebbe ingannarlo in tutti i suoi pensieri e in ogni sua percezione, tutto ciò che lui ritiene esistente potrebbe essere null'altro che il frutto di un'illusione, e quindi di tutto si può dubitare... tranne che di una cosa: del fatto che lui stesso stia dubitando, cioè che stia pensando, e quindi esista. *Cogito ergo sum*. Da questa intuizione (tutt'ora discussa, e la cui validità non manca d'esser messa in dubbio) Cartesio deriva tutta la sua filosofia.

Già gli antichi greci, sin dai tempi di Parmenide, credevano che una realtà *in divenire*, in continuo mutamento così come appare ai sensi, non lasciasse alcuna base per la conoscenza (come, se ciò che io ora vedo tra un secondo sarà diverso da com'era?), solo una realtà stabile nel tempo può essere indagata e conosciuta. Similmente, per Cartesio, negata l'affidabilità dei dati sensoriali, la conoscenza non può più essere esperienziale, deve bensì essere razionale come quella matematica:

il dubitare lo ha portato a distillare un assioma, un principio certo, dal quale poi calcolare, in modo indipendente dai sensi (cioè *a priori*), le verità implicate da questo (Cartesio 1637). Coloro che seguono questo modo di pensare sono *razionalisti*.

Ma i filosofi devono essere in disaccordo su tutto, è il loro lavoro: se ci sono dei razionalisti deve anche esserci chi vi si oppone. Il filosofo inglese (e per un periodo medico senza laurea) John Locke, scrivendo pochi anni dalla dipartita di Cartesio, immaginava la mente come una *tabula rasa*: alla nascita non sappiamo nulla, non c'è alcuna idea nella mente, e non possiamo formarne se non rapportandoci col mondo; detto in modo più esplicito, non c'è nulla nella mente che non sia passato per i sensi, ciò che si può conoscere lo si può conoscere solo partendo dall'esperienza (cioè *a posteriori*) (Locke 1690). Coloro che sostengono questa scuola sono *empiristi*.

A Locke si oppose subito Leibniz, filosofo razionalista di una stravaganza tipica barocca nonché ultimo genio universale (fu uno dei matematici più influenti di tutti i tempi, co-scopritore con Newton del calcolo infinitesimale, inventore del sistema binario, pioniere dell'informatica ante litteram, avvocato, sinologo, ingegnere, filologo, e un mare di altre cose, dando in pressoché ogni campo contributi di rilievo). Non posso negare le mie simpatie per questo grande uomo che credo svettare sopra tutti gli altri; così come non posso negare che, per lungo tempo, la sua filosofia, strabordante di idee tanto originali quanto improbabili, fu più derisa che studiata, presa come esempio di quante assurdità la ragione può produrre senza l'aiuto dell'esperienza: questo è il migliore dei mondi possibili, il mondo è composto da una infinità di sostanze non interagenti, non c'è alcuna interazione tra mente e corpo, e se io mi taglio un dito e sento dolore non è perché una cosa causi l'altra ma solo perché Dio ha prestabilito armonia tra le sostanze... (Leibniz 1710, 1720) chi può crederci seriamente?

Scontro tra titani: Hume e Kant

E mentre Leibniz cadeva in disgrazia, la terra scozzese, da poco entrata nel diciottesimo secolo, dava alla luce David Hume, un filosofo la cui influenza sul mondo anglofono è rimasta senza pari: basti dire che tutt'oggi Hume risulta essere il filosofo con il quale la maggior parte dei filosofi analitici identifica sé e il proprio pensiero (Bourget & Chalmers 2014).

La prospettiva empirista di Hume fu radicale nel mettere in crisi diverse nozioni che fino ad allora si consideravano pressoché indubitabili. La gente dà per scontato che uccidere sia male? Hume tira una linea netta a dividere fatti e valori, mettendo in dubbio la possibilità di fondare un'etica razionale o di pervenire ad una conoscenza del bene e del male. La gente dà per scontato di avere identità personale, di essere un *io*? Hume lo nega, non c'è alcun *io* che fa esperienza del mondo, questo *io* non è altro che l'insieme di tali esperienze. La gente dà per scontato che il calcio causi il volo del pallone? Per Hume non è detto che sia così, magari la prossima volta il pallone non volerà via; io non posso individuare cause ma solo ricorrenze, *di solito* quando calcio il pallone questo vola via, ma per quante volte possa ripetere le osservazioni non potrò mai essere certo della necessità di quello che vedo (Hume 1739, 1748).

È tutt'ora dubbio quali fossero le reali convinzioni di Hume: può sembrare uno scettico tale da dubitare addirittura della sensatezza del concetto di conoscenza, ma pare anche un naturalista fiducioso nella capacità della scienza di dirci qualcosa sul mondo. I problemi che solleva erano e sono reali e difficili da affrontare. Kant, il più grande filosofo moderno, parte dai dubbi di Hume e galoppa come cavaliere dell'intelletto alla difesa della razionalità illuminista.

Kant è conscio dei meriti sia di razionalisti che empiristi, cerca così di arrivare ad una visione media tra i due estremi: riconosce l'esistenza della *cosa-in-sé*, di un mondo esterno alla mente e al pensiero, ma nega che esso sia raggiungibile in modo diretto dall'esperienza, l'esperienza non è pura bensì va interpretata tramite schemi; sì, c'è un universo indipendente da tutto ciò che noi possiamo pensare su di esso, ma non possiamo conoscerlo se non attraverso lo specchio deformante dei concetti che le teorie scientifiche ci mettono a disposizione (Kant 1781). Questa posizione, chiamata *idealismo trascendentale* o *critico*, è paragonata da Kant stesso alla rivoluzione copernicana: non interpreta più la conoscenza come un conformarsi dell'intelletto all'oggetto fisico presente nel mondo, ma al contrario è l'oggetto a conformarsi all'intelletto; non si può conoscere le *cose-in-sé*, ma solo le cose così come appaiono.

Mill e Russell contro l'idealismo tedesco: la prima differenza

Riecco J. S. Mill, anche lui non solo grande filosofo ma anche prodigio della matematica. Ho detto che Kant cercava una via media tra empirismo e razionalismo: come, prima di lui, Aristotele aveva fuso in un unico sistema tutta la filosofia greca, e Tommaso d'Aquino aveva poi fuso Aristotele con tutta la tradizione della patristica, così Kant fonde ora le due grandi tradizioni moderne. Ma Mill non è d'accordo (il principale servizio dei filosofi alla società è il non essere d'accordo con ciò che altri filosofi dicono). Lui, con fare critico, chiama *continentali* proprio quei filosofi che sono influenzati dalle idee di Kant e li contrappone agli empiristi inglesi: non vede in Kant una sintesi, ma una continuazione del razionalismo o al massimo una adulterazione del vero empirismo.

Il sistema kantiano è modesto, arrossisce nel porre una teoria della conoscenza tanto raffinata quanto semplice, ma non fa avventate affermazioni sulla natura del mondo. I suoi discepoli, quelli criticati da Mill, non furono altrettanto modesti...

Un filosofo tutto sommato minore, Schulze, ebbe l'intuizione (enorme, altro che minore) che l'uomo rileva la causalità all'interno dell'esperienza, e se non possiamo conoscere la natura della *cosa-in-sé* non è neanche legittimo ipotizzare che l'esperienza sia *causata* dalla *cosa-in-sé* (Schulze 1792). E allora, non è neanche utile ipotizzare che esista, la *cosa-in-sé*, e le rappresentazioni mentali non sono più rappresentazione di qualcosa, sono esse stesse il mondo. I pensatori successivi dovettero rinunciare a questa *cosa-in-sé*: Fichte (1794) conclude che il soggetto non riceve informazioni dal mondo esterno ma lui stesso, il suo ego, genera le idee; Schelling (1801) invece afferma che non c'è alcuna distinzione tra l'ideale e il reale, tra la *cosa-in-sé* e la cosa così come appare, tanto che ogni cosa può essere considerata identica al suo opposto. Questa linea di pensiero, detta *idealismo tedesco*, incentrata sul primato della mente rispetto al mondo esterno e sulla fusione degli opposti, avrà il suo massimo coronamento in re Hegel, il leone della superbia totalmente immemore della modestia kantiana, autore di un sistema filosofico rococò complesso e, lo posso dire senza troppo timore, incomprensibile.

L'influenza dell'idealismo tedesco, tramite Hegel, fu enorme e diede origine a movimenti idealistici anche in America e Inghilterra. Proprio l'Inghilterra dà i natali ad un altro filosofo-matematico, Bertrand Russell, figlioccio di Mill nonché uomo di ferro morto a 98 anni dopo essere sopravvissuto ad un disastro aereo e a quattro matrimoni (impresa, quest'ultima, ancora più impressionante della prima).

All'epoca della sua giovinezza (nacque nel 1872) l'accademia inglese era dominata dal pensiero hegeliano, e lui stesso vi aderiva; ben presto, però, iniziò a trovare quella filosofia insoddisfacente e

incapace di parlare della realtà: sviluppò così un nuovo metodo filosofico, incentrato sull'analisi del linguaggio tramite gli strumenti della logica matematica. Russell, in modo simile a Mill, distinse i filosofi *analitici* dell'accademia inglese dai *continentali* degli altri Stati europei, cioè coloro che erano ancora influenzati dall'idealismo tedesco (Russell 2004). È un'affermazione particolarmente autorevole visto che proprio lui, Russell, può essere considerato il padre della filosofia analitica.

Ecco quindi una prima differenza storica tra la tradizione analitica e quella continentale: mentre i continentali hanno accettato la sintesi kantiana e l'idealismo che ne è derivato, gli analitici si sono mantenuti fedeli alla tradizione dell'empirismo iniziata con Locke, rifiutando prima il razionalismo cartesiano, poi l'idealismo trascendentale, e infine l'idealismo tedesco.

La scuola di Bolzano e Brentano

Continuo con i filosofi-matematici: Bernard Bolzano, che scriveva «da mia speciale predilezione per la matematica si fonda in modo particolare sui suoi aspetti speculativi, in altre parole apprezzo molto quella parte della matematica che è stata allo stesso tempo filosofia» (Valerio 2017). È proprio questa unità tra le due sfere della ricerca intellettuale, la creatività e il rigore, che differenzia Bolzano dagli idealisti suoi contemporanei.

Il suo lavoro filosofico è tanto ordinato e razionale quanto quello matematico. Nel tentativo di dare un fondamento logico a tutte le scienze crea un sistema di pensiero ricco di astrazioni quali proposizioni, relazioni, parti, idee, forme, insiemi, sostanze, giudizi (Bolzano 1837). La sua analisi di questi concetti è sempre rigorosa, spesso formalizzabile, traducibile nel linguaggio della logica matematica; esplora il significato di tali termini procedendo per esperimenti, mostra come il loro significato cambia all'interno di diverse espressioni; struttura i suoi ragionamenti per proposizioni, espone ordinatamente le premesse e le regole che usa per derivare da queste premesse le sue conclusioni proprio come si faceva nei trattati di geometria, senza affidarsi a intuizioni vaghe o concetti oscuri. La filosofia di Bolzano, proprio come quella di Russell, va giocata tra linguistica e matematica.

Il maggior erede della scuola di Bolzano fu Franz Brentano, prete cattolico boemo come lo stesso Bolzano ma anche pioniere di quella scienza all'epoca neonata che è la psicologia. Ebbe grandiosa influenza su molti suoi allievi, tra i quali ricordiamo almeno Freud (proprio lui) e Alexius Meinong (stravagante filosofo, considerato italiano dagli austriaci e austriaco dagli italiani, famoso per la sua *giungla*, il territorio metafisico che contiene gli oggetti che sono reali ma non-esistenti) (Meinong 2019). Il contributo più importante di Brentano fu il recupero del concetto di *intenzionalità*, capi, cioè, che la caratteristica fondamentale della vita psichica sta nel suo essere diretta verso qualcos'altro: io non desidero e basta, ma desidero qualcosa; non penso e basta, penso a qualcosa; non credo e basta, ma credo qualcosa (Brentano 1874).

Partendo da questo concetto, il più grande allievo di Brentano, Edmund Husserl, iniziò a concepire un nuovo tipo di filosofia.

Matematica, psicologia e filosofia tra Husserl e Frege: la seconda differenza

Husserl, forte della sua educazione scientifica, scende nel campo della filosofia rimanendo con un piede nella matematica e l'altro nella psicologia: il suo obiettivo era infatti quello di dare un fondamento psicologico (ma non *psicologista*, non *riduzionista*) alla matematica, tentava cioè di mostrare come la mente arrivasse a concepire per la prima volta il numero e tutto ciò che vi consegue. Questo obiettivo era lo stesso che si era posto il matematico suo contemporaneo Gottlob Frege.

Il rapporto tra Husserl e Frege è complesso, capirlo vuol dire capire anche il generale rapporto della filosofia analitica con la continentale. È opinione diffusa che Frege sia il nonno della filosofia analitica, mentre Husserl sia un esponente della filosofia continentale: qui nel Continente in pochi studiano Frege, ma l'importanza di questo pachiderma filosofico nella tradizione analitica è così grande che non è mancato chi lo ha paragonato addirittura ad Aristotele. Eppure Husserl e Frege si conoscevano, si scrivevano, uno capiva il lavoro dell'altro, e con ogni probabilità non si erano mai percepiti come appartenenti a due tradizioni diverse, nonostante l'abissale diversità dei loro lavori anche da un punto di vista stilistico (Dummett 2001).

Il pensiero maturo di Husserl sposta il suo focus dai fondamenti della matematica alle *strutture della coscienza* in generale: al modo degli idealisti, sceglie di studiare gli oggetti non come cose esterne ma come insiemi di percezioni e funzioni che si danno all'interno dell'esperienza, il rapporto dell'uomo con essi è descritto dall'intenzionalità; non riduce la coscienza ad altri elementi del mondo, ma al contrario interpreta il mondo tramite l'apparenza cosciente, sospende il giudizio sulla natura della *cosa-in-sé* (Husserl 1900). È il tentativo di un discorso oggettivo e scientifico (in senso ampio) sulla soggettività, portato avanti non più raccogliendo dati empirici come fa la psicologia ma parlando direttamente dell'esperienza.

La riflessione di Husserl ha originato non solo una tradizione ma addirittura un settore della filosofia, un nuovo insieme di domande e problemi, un settore che Husserl ha considerato essere la *filosofia prima*, quel settore che sta alla base di tutti gli altri e offre i concetti basilari necessari a rispondere a tutte le altre domande. Il nome di questa tradizione e di questo settore è *fenomenologia*, lo studio della coscienza in sé stessa.

Il miglior allievo di Husserl, Heidegger, ha usato la prospettiva fenomenologica per affrontare alcuni dei più profondi problemi metafisici: ha interpretato la coscienza come un effetto dell'esistenza per poi studiare l'esperienza dell'azione e del significato (Heidegger 1927); la sua è una delle filosofie più oscure e influenti di tutti i tempi. Ma anche molti altri hanno seguito le orme di Husserl fondendo la fenomenologia con pressoché tutte le altre scuole di pensiero. Una particolare tradizione etica interna alla fenomenologia è quella del *personalismo*, una scuola filosofica che ha avuto tra i suoi esponenti un personaggio famosissimo anche tra coloro che non hanno mai letto alcunché di filosofia: Giovanni Paolo II.

Frege, pur diretto verso gli stessi obiettivi di Husserl, ha sviluppato il suo lavoro in tutt'altra direzione. Rifiutando, come Husserl, lo *psicologismo* (l'idea che la matematica sia riducibile alla psicologia) Frege porta avanti la tesi che, come la quercia dal seme, la matematica germogli tutta dalla logica. Per tentare di dimostrarla Frege ha dovuto creare un nuovo settore della matematica, un nuovo tipo di logica più versatile di quello che esisteva alla sua epoca e capace di rappresentare formalmente un gran numero di espressioni prima non formalizzabili (Frege 2021). La potenza di questo nuovo mezzo formale ha dato a Frege la possibilità di analizzare con rigore mai visto prima tematiche inerenti la linguistica, un campo nel quale ha distinto, tra le componenti del *significato*, il

referimento (l'oggetto concreto a cui un nome si riferisce) e il *sensò* (il suo contenuto espressivo-concettuale) (Frege 2007).

L'intuizione di Frege ha effetti enormi: sostituendo, nella ricerca dei fondamenti della matematica, l'analisi psicologica con la logica legittima l'uso di formalismi come mezzo per rappresentare le strutture della mente. In Frege le parole non hanno un senso *indagabile* se non nell'insieme del senso di una frase, non posso indagare il concetto di numero al di fuori delle proposizioni che contengono un numero (lo chiama *principio del contesto*): quindi non c'è neanche analisi del pensiero che non sia analisi delle proposizioni. Questa mossa lo differenzia da Husserl come il giorno dalla notte: mentre Husserl tenta un approccio diretto, per Frege gli stati mentali sono indagabili solo tramite il linguaggio e la logica (io penso *a qualcosa*, in modo intenzionale, ma questo *qualcosa* è espresso da una proposizione, a sua volta esprimibile tramite una frase del linguaggio, e quindi analizzabile tramite mezzi formali).

Husserl e Frege mostrano due modi diversi di intendere la mente, una diversità che costituisce un secondo punto di divisione tra filosofia continentale e analitica: mentre i continentali, seguendo la scia di Brentano e Husserl, esplorano la mente in modo frontale appellandosi direttamente ai dati esperienziali, gli analitici, seguendo Bolzano e Frege, credono che il linguaggio non segua ed esprima il pensiero bensì lo preceda e lo renda possibile, e danno così al linguaggio e alla logica un primato come mezzi d'eccellenza tramite i quali comprendere la mente umana.

Le proposizioni analitiche: la terza differenza

La terza differenza tra filosofia analitica e continentale è immediata.

Kant (1781) divide le proposizioni in due tipi: *analitiche* e *sintetiche*. Scrivo una proposizione qualsiasi « x è y » dove x è il soggetto e y il predicato. Le proposizioni analitiche sono quelle vere in virtù del loro stesso significato in quanto la predicazione è contenuta nel soggetto, mentre le proposizioni sintetiche richiedono dati ulteriori al loro significato in quanto la predicazione non è contenuta nel soggetto.

Per esempio, «nessuno scapolo è sposato» esprime una proposizione analitica (perché il non essere sposati è la definizione di scapolo), mentre «tutti gli uccelli hanno un intestino» ne esprime una sintetica (perché l'aver l'intestino non è la definizione di uccello). La verità di una proposizione analitica può essere conosciuta *a priori*, tramite la sola ragione in modo indipendente dall'esperienza, mentre una affermazione sintetica va conosciuta *a posteriori*, tramite indagine sul mondo. Secondo Kant esistono inoltre proposizioni *sintetiche a priori*, come quelle della matematica.

Il processo di analisi consiste nella scomposizione di un concetto complesso nelle sue componenti concettuali più semplici: per esempio, un'analisi di *scapolo* è *uomo* e *non-sposato*, l'analisi produce un suo sinonimo esteso.

Frege voleva analizzare la matematica tramite la logica, cioè dimostrare che è possibile formare espressioni analitiche dove il soggetto è una frase della matematica e la predicazione una della logica: così, ogni espressione matematica ha un sinonimo in logica, ed è possibile tradurre la lingua della matematica nella lingua della logica (Frege 2019).

Vi era la speranza che, alla lunga, si sarebbe colta la necessità delle verità matematiche, e che questo fosse poi estendibile alla scienza e a una pluralità di concetti filosofici. In che modo? La verità di una frase come «nessuno scapolo è sposato» è analiticamente evidente in quanto sostituendo a «scapolo» la sua analisi si ottiene «nessun uomo non-sposato è sposato»; così, l'obiettivo (o speranza) era mostrare che molte domande filosofiche sull'esistenza, la conoscenza, il bene fossero domande come «ogni medico è anche un dottore?», pseudo-dilemmi dissolvibili tramite la sola analisi del significato delle parole.

Questo è il motivo per cui la filosofia analitica si chiama così, è la sua essenza storica ed è qualcosa di alieno alla mentalità continentale.

Così, confrontando la filosofia analitica con l'idealismo ho già individuato tre differenze; per controllare se ve ne sono altre, devo ora parlare delle altre scuole continentali.

Ermeneutica ed esistenzialismo

Come la fenomenologia, anche l'*ermeneutica* è sia una scuola che un settore della filosofia, è il settore che si occupa delle domande relative all'interpretazione nel senso più generale (non solo come si interpreta un testo, ma anche un oggetto, un atteggiamento, qualsiasi cosa possa trasmettere significato). E come la fenomenologia, anche l'*ermeneutica*, in virtù della versatilità dei suoi strumenti, è stata fusa con pressoché tutte le altre scuole: *ermeneutica* e fenomenologia davvero vanno a braccetto, se descrivo il modo in cui gli oggetti mi appaiono descrivo anche il modo in cui il loro significato sorge nella mia esperienza. Questo si vede bene in Heidegger (1927), considerato contemporaneamente tra i maggiori esponenti della fenomenologia e dell'*ermeneutica*: usa quest'ultima per parlare del significato da un punto di vista esistenziale e del modo in cui l'uomo comprende l'esistenza degli altri individui (la sua filosofia viene chiamata *fenomenologia ermeneutica trascendentale*).

Questo mi porta all'*esistenzialismo*, le cui acque sono molto più impervie di quanto sembri.

È un movimento ambiguo: secondo alcuni non è neanche una scuola filosofica, è un clima culturale che ha legami più profondi con la letteratura che con la riflessione teoretica. I suoi esponenti sono più famosi come scrittori o aforisti che come filosofi, sono tra i pochi personaggi citati in queste pagine che possono essere letti anche solo per piacere: ecco Dostoevskij, la vetta ghiacciata della letteratura russa; Kierkegaard, il pensatore danese più ricco di Batman che ha passato tutta la sua esistenza chiuso in un castello; Nietzsche, il Thor della filosofia, che con il suo martello saettante ha fatto a pezzi, un aforisma alla volta, il razionalismo moderno da lui odiato. Significativo che Russell (2004) descriva la filosofia continentale come *letteraria*, una caratteristica comune dell'*esistenzialismo* e del pensiero francese (i grandi filosofi francesi, da Voltaire e Rousseau fino a Camus, hanno espresso il loro pensiero tramite poemi, racconti, romanzi, più che saggi).

È difficile trovare elementi unificanti tra i vari pensatori che, quasi tutti a posteriori, sono stati indicati come esistenzialisti, così come è difficile dire se oggi esistano degli esistenzialisti e non sia invece da considerare un movimento storico. L'*esistenzialismo* sceglie, al modo della fenomenologia, di partire dal soggetto esperente, e in particolar modo dalla sua esperienza di ansia e terrore davanti ad un mondo che pare assurdo, assurdo in quanto indeterminato, libero, che lascia sulle spalle dell'individuo tutto il peso delle sue scelte destinate ad apparire in ultimo luogo

immotivabili (Sartre 1946). E questo è il collegamento con Heidegger, cavaliere di gran croce decorato di gran cordone della fenomenologia, dell'ermeneutica, e ora anche dell'esistenzialismo: lui infatti è la principale ispirazione dietro il lavoro di Sartre, il primo scrittore a chiamarsi esistenzialista.

Heidegger ha, nella tradizione continentale, lo stesso ruolo che San Tommaso ha nel pensiero cristiano: è il *doctor communis*, il pensatore che nessuno dei suoi successori, qualsiasi cosa voglia scrivere, può permettersi di ignorare. Nella tradizione analitica, invece, è forse il pensatore più disprezzato: Russell lo accusa di oscurantismo (2012); A. J. Ayer, il peso massimo della razionalità scientifica del Novecento, considera la filosofia heideggeriana come inutile e dannosa in quanto inverificabile tramite ricerca empirica; Carnap, altro gigante tra i filosofi analitici, sviluppa una analisi in logica formale di varie affermazioni di Heidegger dimostrando che violano regole di sintassi e che per questo risultano insensate (Ayer 1959).

Tra continentali e analitici c'è una notevole differenza di canone: se nel pensiero continentale Hegel ed Heidegger, pur in contrasto, sono i due giganti indiscutibili degli ultimi due secoli, gli analitici tendono a considerarli quasi dei minori; questo, però, è già implicito nel rifiuto dell'idealismo.

Il libro si tinge di rosso: il Marxismo

Come l'esistenzialismo, anche il *marxismo* può non sembrare una filosofia (e questa è una cosa della quale i marxisti sono ben contenti); forse non lo è, forse è qualcosa in più di una filosofia, è una teoria politica, economica e sociologica. L'influsso del marxismo su pressoché tutti i campi del sapere umanistico (dalla critica letteraria all'archeologia, dalla geografia alla pedagogia, dalla criminologia alla psicoterapia) è enorme, tanto da essere stato uno dei motori dietro i più importanti eventi storici degli scorsi due secoli (questo lo dico a quelli che pensano che la filosofia non faccia nulla di concreto: a volte causa addirittura milioni di morti...).

Il marxismo si basa su due presupposti. Il primo è il *materialismo storico*, l'idea che il susseguirsi degli eventi storici sia guidato non dagli ideali, ma dalle condizioni materiali della società che si studia, cioè dai fattori economici e tecnologici che motivano gli ideali. Il secondo è il *materialismo dialettico*, una interpretazione del mondo naturale che passa non attraverso le percezioni o le teorie scientifiche, bensì attraverso le interazioni socioeconomiche tra le varie classi sociali, soprattutto tra i lavoratori e i datori di lavoro.

Marx pensava che la società del suo tempo fosse ferita da un contrasto non sostenibile tra il rapido sviluppo tecnologico della rivoluzione industriale e gli altri fattori sociali, filosofici, artistici che sembravano essere rimasti all'epoca precedente. Quando si creano questi contrasti ne risultano incoerenze che si manifestano nel conflitto tra le classi sociali; l'unico modo per rendere la struttura sociale coerente con l'attuale stato tecnologico è una rivoluzione che deve ribaltare la piramide sociale, dare potere ai lavoratori e, alla lunga, portare a non avere più bisogno di uno Stato (Marx 1867). Non si può fare il contrario, modificare la tecnologia, distruggerla se serve, per renderla coerente con la struttura sociale esistente? Questo è il *luddismo*, che nel sistema marxista è impossibile.

Mi sembra di risentire le frasi di quel breve testo, *il più antico programma sistematico dell'idealismo tedesco* (non si sa se scritto da Hegel, dal suo compagno di stanza Schelling, o dal grande poeta Holderlin)

che molti decenni prima di Marx affermava «soltanto ciò che è oggetto di libertà può chiamarsi idea. Dobbiamo dunque superare lo Stato! Ogni Stato, infatti, necessariamente considera gli uomini liberi come ingranaggi meccanici, ma così non deve essere, e lo Stato deve dunque estinguersi.» (Holderlin 2004)

Anche se materialismo storico e dialettico sono accettati da tutte le scuole marxiste ortodosse, i dettagli della filosofia di Marx sono stati rivisti, reinterpretati e trasformati in ogni modo: oggi il marxismo non è più una scuola di pensiero, è diventato un insieme di scuole con più capostipiti. Questa infedeltà nei confronti di Marx è stata resa necessaria dagli sviluppi della teoria economica dell'ultimo secolo e dall'applicazione sempre più rigorosa di modelli matematici e simulazioni a computer di problemi socioeconomici: bisogna infatti riconoscere che almeno alcuni elementi della teoria marxista sono oggettivamente errati (Howard & King 1992). Il filosofo della scienza Popper, analitico, notava che, per quanto la teoria di Marx avesse ambizioni scientifiche e ponesse delle previsioni sugli effetti delle manovre economiche che potevano essere verificate tramite esperimenti empirici, essa è stata sfatata proprio da tali test (Popper 2004); Russell, più duro, bolla la teoria marxista direttamente come pseudoscienza (Russell 2004).

Teoria critica e strutturalismo

La visione di Marx porta ad interpretare tutte le ideologie politiche, i movimenti artistici, le filosofie, come discendenti da un certo stato materiale, in particolar modo dal metodo di produzione dei beni, dalla struttura del lavoro dell'epoca di riferimento. È facile immaginare l'effetto che questa idea ha in ogni studio di tipo umanistico: la letteratura, la filosofia, le scienze politiche, ogni materia va rivisitata, non deve più essere interpretata come qualcosa che può determinare il corso della Storia ma al contrario va studiata come prodotta dalla Storia. È un rovesciamento nell'interpretazione del mondo oggettivamente geniale.

Marx credeva che l'alienazione, il sessismo, il diverso livello di benessere tra città e campagna e tutto ciò che c'è di male nel mondo fosse effetto del capitalismo, e lamentava che le filosofie del passato, pur potendo evidenziare il problema, non avessero mai cercato di fare alcunché per cambiare la situazione. Colto questo richiamo, i filosofi della *teoria critica* si appoggiano alle scienze sociali per esprimere una critica o analisi della cultura con l'obiettivo di cambiarla, cercano di capire come cancellare il peso delle ideologie passate e di ogni forma di autorità: non è una teoria in senso solo descrittivo, ma anche attivo, non vuole solo rappresentare il mondo ma anche agire politicamente su di esso.

Le specifiche idee della teoria critica, come nel caso del marxismo, non sono esplicitabili: nella sua storia ha subito molte trasformazioni, man mano che le previsioni di Marx si dimostravano fallimentari ha dovuto evolversi, è avanzata per periodi sostenendo di volta in volta tesi diverse, anche contraddittorie, che hanno negato vari aspetti dell'ortodossia marxista.

La teoria critica tende oggi a sovrapporsi con un'altra scuola, quella del *post-strutturalismo*.

Lo *strutturalismo* costituiva una teoria generale in sociologia, antropologia, e linguistica secondo la quale la cultura umana va studiata come una struttura, cioè i suoi elementi non sono comprensibili se non nell'insieme, considerando il modo in cui i vari aspetti delle credenze, delle strutture sociali, e soprattutto del linguaggio si relazionano tra loro (Saussure 1916). Questo *olismo* era già presente, in nuce, nell'idealismo tedesco (Hegel credeva che vi fossero cose definite dalle loro relazioni con il

resto del mondo, e quindi che non fossero conoscibili se non conoscendo prima tutto il resto), e mantiene collegamenti con l'ermeneutica e la fenomenologia (la conoscenza della struttura permette di conoscere anche l'esperienza e capire la base del significato che diamo alle cose), ma l'importanza che dà al linguaggio è coerente anche con la tradizione analitica.

Ben presto, vari sostenitori dello strutturalismo iniziarono a perdere fiducia nella possibilità di comprendere alcunché in questo modo: in fondo, i filosofi stessi agiscono dall'interno di una struttura, loro stessi appartengono ad una cultura che forma il modo in cui vedono il mondo, e non sembra esserci modo per rendersi indipendenti da queste influenze. Lungi dall'essere interpretato come un paradosso, questo è l'atto di nascita del post-strutturalismo: non è un *anti-strutturalismo*, ma quasi un *super-strutturalismo*, le premesse dello strutturalismo sono estremizzate al punto che finiscono per auto-rifiutarsi.

Un tipo particolare di post-strutturalismo è il *decostruzionismo* di Derrida, una filosofia incentrata sull'analisi delle parole (in particolare quelle esprimenti valori centrali nella cultura occidentale) tramite il loro confronto con le altre parole del linguaggio e con i vari stati passati del linguaggio (Derrida 1967); anche questo approccio affonda le sue radici in Heidegger.

L'autorità della scienza: la quarta differenza

Sia nel post-strutturalismo che nella teoria critica c'è una perdita di fiducia nella libertà umana: l'uomo agisce dall'interno di una certa cultura e non ha modo di uscirne o rendersi indipendente. Da questo si deriva una relativizzazione delle conoscenze: i pensatori di queste scuole sono felici di ridurre l'autorità delle loro stesse teorie, usano le conoscenze delle scienze sociali proprio per dimostrare che esse non sono in grado di produrre conoscenza oggettiva, un limite che non va corretto ma accettato ed esaltato come liberante e democratico. Tale atteggiamento si estende in un generale scetticismo verso le scienze naturali, che vengono viste come pratiche culturalmente limitate e influenzate dai fattori sociali e politici tanto quanto qualsiasi altro tipo di ideologia (Feyerabend 2013). È un ribaltamento del progetto kantiano: mentre la critica di Kant è un'epurazione da ciò che è indimostrabile per salvare la razionalità e il realismo dallo scetticismo di Hume, qui la critica finisce per giustificare uno scetticismo ancora più radicale.

Michel Foucault, secondo i dati di Google Scholar, è l'autore più citato e influente di sempre: il suo h-index (che rappresenta la produttività e il numero di citazioni) è quasi 300 (per fare un paragone, l'h-index medio di un premio Nobel è inferiore a 70): posso ben riconoscerlo come uno dei più grandi intellettuali di sempre. Foucault, *archeologo delle idee* e maggior rappresentante della tradizione continentale dopo Heidegger, nel 1971 partecipò con Noam Chomsky, per acclamazione universale il più grande linguista di tutti i tempi ed emblema del pensiero analitico, ad un dibattito sulla natura umana: Chomsky sosteneva che essa consiste in strutture innate del cervello, Foucault invece ha portato avanti una visione sociologica relativistica (Chomsky & Foucault 2005). Personalmente, ho visto questo dibattito una sola volta, e non voglio rivederlo (non mi piacciono le carneficine, un continentale non può aver ragione su di un vero scienziato), ma è diventato nella mia fantasia il massimo simbolo del conflitto tra queste due visioni del mondo, tra l'idea che i problemi scientifici siano storici e transculturali, e l'idea che invece la scienza dipenda da concezioni preteoriche guidate da influssi sociali e storici modificabili e che le varie teorie siano accettate o respinte a seconda degli elementi culturali tra i quali la teoria cerca di inserirsi.

Negli anni Novanta questo conflitto ha raggiunto il parossismo con le *science wars*, una serie di dibattiti feroci tra queste due tradizioni. La ciliegina sulla torta fu il celebre *affare Sokal*: il fisico Alan Sokal inviò ad alla rivista continentale *Social Text* una ricerca, scritta nello stile tipico della filosofia post-strutturalista ma insensata e priva di qualsiasi sforzo intellettuale, nella quale sosteneva che tutto il campo della gravità quantistica fosse un costrutto sociale... e ottenne la pubblicazione (Sokal 1996). Questa fu per molti la dimostrazione che la filosofia anti-scientifica di stampo continentale era approssimativa se non del tutto insensata. Forte di questa esperienza, in Sokal 1999 l'autore analizza vari passi falsi dei filosofi continentali in filosofia della scienza, come ad esempio l'idea della femminista, linguista e psicologa Luce Irigaray che $E = mc^2$ sia un'espressione sessualizzata perché privilegia la velocità della luce su altre velocità per noi vitali, o che la meccanica dei fluidi sia stata trascurata perché femminile rispetto alla più mascolina meccanica dei solidi (Irigaray 1978, ch. 6).

In effetti, ho motivi per pensare che la diffidenza nei confronti della conoscenza scientifica sia il centro del contrasto tra la tradizione analitica e quella continentale: questa è la quarta grande differenza. Scorrendo i dati di Bourget & Chalmers 2014 si nota che la correlazione tra il dichiararsi appartenenti alla tradizione continentale e il dichiararsi scettici nei confronti della scienza è forte quasi quanto quella che c'è tra l'occuparsi di filosofia della religione e il credere in Dio: potrei dire che essere diffidenti verso la scienza è quasi un prerequisito per i filosofi continentali. Una straripante maggioranza di filosofi analitici, al contrario, accetta i dati che arrivano dalle scienze naturali come le migliori conoscenze desiderabili, e non si sognano neanche di notte che una teoria filosofica possa rivedere e modificare la visione scientifica del mondo.

Né analitici né continentali: le altre scuole

Confrontando analitici e continentali ho individuato quattro differenze principali. Nella nebbia di guerra tra i due fronti, però, si nascondono altre tradizioni filosofiche, non riducibili ad uno dei due schieramenti, e dalle quali la filosofia analitica si distingue. Devo quindi dare un'occhiata a queste piccole tradizioni, e controllare se ci sono altre differenze di rilievo o punti di contatto.

Riprenderò spesso il tema del contrasto tra essere e divenire: io sono io, eppure io sono diverso da com'ero dieci anni fa; com'è possibile accodare le due cose, la mia identità con il fatto evidente che cambio di continuo? Parmenide risponde che i sensi mi ingannano, che il divenire è illusorio, e c'è solo l'essere: questa filosofia, che definisco *basata sulla sostanza*, è stata anche sposata e sviluppata dal più celebre filosofo italiano dei tempi recenti, Emanuele Severino (che pare essere morto, ma ciò sarebbe in contrasto con la sua filosofia, quindi non si sa). Eraclito, al contrario, afferma che è l'idea di un essere eterno e stabile ad essere illusoria, e che tutto è in divenire, tanto che un uomo non può mettere due volte i piedi nello stesso fiume: questa è una filosofia *basata sul processo* (Diels & Kranz 1903).

Duemila anni dopo Parmenide ecco Alfred North Whitehead, il maestro di Russell. Inizialmente dedito solo alla matematica, col tempo i suoi interessi si spostano verso la metafisica e il rapporto tra scienza e filosofia: inizia a scrivere libroni oscuri come le citazioni di Eraclito, ricchi di una filosofia originale e complessa. Il suo pensiero, la *filosofia del processo*, vuole revisionare la filosofia naturalistica sostituendo il concetto di sostanza, che domina il pensiero occidentale, con quello storicamente più sfortunato di processo, in modo da creare una metafisica alternativa che possa rendere meglio merito dei dati delle scienze, una metafisica dove gli enti non sono oggetti con una essenza stabile ma sono definiti dal loro stesso mutamento (Whitehead 1929).

Alcuni hanno interpretato il passaggio di Whitehead dalla discussione logico-matematica alla metafisica come un passaggio dalla filosofia analitica a quella continentale, ma la filosofia del processo è così particolare da doversi considerare una tradizione a sé, con metodi e temi che sono a mezza via tra le due tradizioni maggiori. Di fatto, si tratta dell'ultimo sistema filosofico mai prodotto, l'ultimo tentativo di sviluppare una teoria generale sul mondo (questo, invero, non è stato fatto da Whitehead, ma dal piccolo manipolo dei suoi seguaci). Oggi la filosofia del processo è stata accolta soprattutto da quei filosofi che hanno un interesse per l'ecologia, ma ha influenzato anche teologi e psicologi.

Un'altra tradizione con radici profonde è quella del *tomismo*, la filosofia di San Tommaso. È di gran lunga la tradizione filosofica più antica tra quelle ancora attive: dopo più di settecento anni non ha ancora smesso di produrre riflessioni (le altre tradizioni tutt'ora attive non hanno più di duecento anni). È però una delle filosofie più rigide visto che ha ormai una sua risposta ortodossa per tutte le questioni filosofiche che sono emerse lungo i secoli. È una filosofia tecnica, incentrata sulla logica, quasi come una filosofia analitica ante-litteram (mentre la filosofia continentale può risuonare con la filosofia rinascimentale fondata sulla filologia e la letteratura classica). Si tratta, infine, dell'ultimo tentativo di fondere in un unico sistema coerente tutta la riflessione prodotta da tutti i filosofi precedenti: se Aristotele è il compimento e la pienezza e il travalicamento di tutto il pensiero greco, la filosofia di Tommaso è il riflesso di una luna aristotelica nel lago della patristica.

L'incapacità del progetto illuminista di produrre un'etica razionale e dare un senso alla vita che sia percepito come tale ha portato oggi molti filosofi, anche atei, a guardarsi indietro, a risalire a Tommaso; alcuni, ispirati dal suo progetto di sintesi, hanno fuso il sistema tomista con pressoché ogni altra scuola contemporanea: i gesuiti Bernard Lonergan e Karl Rahner vi hanno aggiunto la filosofia kantiana e l'idealismo tedesco; i santi Edith Stein e Giovanni Paolo II l'hanno farcito con la fenomenologia; il domenicano Bochenski, l'allieva prediletta di Wittgenstein Elizabeth Anscombe, suo marito Peter Geach e vari altri filosofi anglofoni come Edward Feser hanno donato al tomismo il rigore della logica formale e del ragionamento analitico.

Ora due parole sull'*oggettivismo* della scrittrice Ayn Rand. Rand sviluppa il suo pensiero in modo sistematico, va a toccare tutti i grandi temi della filosofia; la sua metafisica e la sua epistemologia, però, sono troppo semplici, fa poco più dell'affermare il primato della ragione, tanto che i tre quarti della sua opera sembrano essere solo una scusa per arrivare alla fine a presentare una dottrina etica e politica originale ma estrema, una forma di individualismo che giustifica forme di anarcocapitalismo o miniarchismo, una sorta di anti-marxismo in cui lo Stato scompare in favore non dei lavoratori ma della libera azione degli imprenditori.

Alla tradizione filosofica americana per eccellenza, quella del *pragmatismo*, fondata da C. S. Peirce, voglio dedicare più spazio, anche per l'affetto che provo verso questo filosofo (Peirce e Putnam nel suo periodo kantiano sono le mie maggiori influenze).

I contributi di Peirce sono impressionanti per ampiezza e originalità: è stato uno dei più grandi logici matematici di sempre, ha creato il settore della semiotica (lo studio dell'interpretazione dei segni), e ha toccato tutti i maggiori settori della filosofia; successi, questi, che sono da considerarsi ancora più incredibili alla luce del fatto che Peirce non fu neanche un filosofo (se non per pochi anni) bensì passò buona parte della sua vita lavorando come scienziato per poi morire in povertà.

Un pragmatista è tale se applica la massima pragmatica. Peirce prova ad esprimere questa massima sette volte nei suoi scritti, e il tentativo meglio riuscito è forse il secondo: «l'intero significato intellettuale di un simbolo consiste nella totalità dei modi generali di condotta razionale che

seguirebbe dall'accettazione del simbolo in tutte le possibili differenti circostanze e desideri»; Peirce infatti credeva, in un primo momento (cambierà idea negli ultimi anni), che «l'intera funzione del pensiero è produrre abiti d'azione», «se qualcuna delle nostre sensazioni non si riferisse a come agiremmo in una determinata occasione... allora non la chiameremo pensare» (Perice 2011). Non intende dire che l'azione è in sé lo scopo e il significato di ogni concezione, ma che ogni concezione risulta in conseguenze: per esempio, il mio concetto di *durezza* mi fa sorgere delle aspettative, mi fa pensare che se andrò a premere l'oggetto duro esso non cambierà forma, e finché non premo, finché non capisco il modo in cui la *durezza* si manifesta nella mia vita, non posso capire la differenza tra duro e morbido.

Peirce definisce il pragmatismo come un metodo di riflessione, senza dottrine specifiche. La massima pragmatica punta al chiarimento dei concetti, non alla verità: dice che il significato del concetto sta nei suoi effetti, che il significato va indagato considerando le conseguenze pratiche discendenti dal considerare quel concetto vero, ma non dice se è in sé vero. La concezione pragmatica della conoscenza è di tipo evolutivo, non è qualcosa che riguarda l'individuo quanto la comunità di coloro che ricercano la conoscenza, e se c'è della verità essa sta nello stato finale ideale, infinitamente futuro, di questa ricerca.

Oggi la tradizione del pragmatismo si è sviluppata fino ad avere una sua opinione distintiva in tutte le maggiori questioni filosofiche; il suo fulcro rimane l'idea che si stia dando all'epistemologia un'importanza eccessiva, l'uomo non è fatto per conoscere (ogni sforzo conoscitivo potrebbe fallire, potrebbe esserci molto di inconoscibile), ma bensì per agire, fare previsioni e risolvere problemi, e farebbe bene a prendere per vero ciò che è utile a tale scopo.

Wittgenstein e i neopragmatisti: forse la filosofia analitica non è ciò che sembra...

Orduque, riassumendo le differenze principali tra analitici e continentali:

1. i continentali danno una importanza capitale all'idealismo tedesco di Hegel e alla filosofia di Heidegger, mentre gli analitici rifiutano questi pensatori;
2. i continentali partono da un accesso diretto alla mente, mentre per gli analitici gli stati mentali sono comprensibili solo tramite la mediazione del linguaggio;
3. i continentali non hanno un metodo specifico e lavorano confrontando i concetti tra loro e ripercorrendone la Storia, mentre gli analitici hanno l'ambizione di risolvere i problemi filosofici tramite un'analisi concettuale lenta e molto specifica che porti a ridurli ad enunciati analitici;
4. i continentali usano gli strumenti della sociologia e della letteratura guardando con scetticismo alle scienze naturali, mentre gli analitici si affidano alla matematica e lavorano in continuità con le scienze.

Mh, non sono proprio sicuro di quello che sto dicendo...

Parto, rifiutando ogni stravaganza, dal primo punto. Se gli analitici respingono Heidegger, chi mettono al suo posto come maggior filosofo del Novecento? Be', Wittgenstein, che io riconosco come secondogenito di Dio.

Inizialmente destinato all'ingegneria (sin da bambino si divertiva a costruire macchine da cucine con pezzi di legno), poi ossessionato dalla filosofia matematica di Russell, radicale ed esplosivo in

ogni sua scelta, pieno di fobie e ossessioni, era dotato di un enorme carisma estetico che emerge potente in tutti i suoi scritti sia in quanto a idee che a stile. È il pensatore più pericoloso, è l'involontario dittatore della filosofia, perché dopo che lo si è letto si cade sotto un'ipnosi e non si può pensare altro che, se esiste una qualche verità a questo mondo, allora deve trovarsi nei suoi scritti.

Inizialmente il *primo* Wittgenstein, quello giovane del *Tractatus* (2009), sposò l'atomismo logico di Russell (2003): il mondo è costituito da fatti che sono rappresentabili dalla logica sviluppata da Frege, per creare un ritratto del mondo basta individuare i fatti atomici e collegarli tramite gli operatori logici; i problemi filosofici nascono dal fatto che il linguaggio normalmente utilizzato per parlare non è adatto a rappresentare il mondo, anzi crea confusione, ed è quindi necessario sviluppare un nuovo tipo di *linguaggio ideale* adatto alla discussione filosofica. I filosofi del *Circolo di Vienna* estremizzarono queste idee nel *positivismo logico*: gli assiomi della logica e della matematica sono tautologie (proposizioni sempre vere, come «una mela o è rossa o non lo è»), le affermazioni delle scienze sono empiricamente verificabili, mentre tutto il resto, tutte le altre affermazioni non verificabili, anche tutte le proposizioni etiche, sono insensate.

Ma questo non è ciò che il criptico Wittgenstein voleva dire. Non aveva mai voluto affermare che le proposizioni etiche sono insensate, al contrario voleva scrivere un libro sull'etica, pur ritenendo di non poterne parlare direttamente. Il suo stile di scrittura ambiguo e aforismatico ricorda Nietzsche o Pascal e non si conforma alla chiarezza scientifica desiderata da Russell e dal positivismo logico. Wittgenstein rifiuta i pensatori ammirati da questa tradizione, si rifà invece a teologi e scrittori: segue sant'Agostino, Kierkegaard, Dostoevskij, Tolstoj, gli aforisti e satiristi Lichtenberg e Kraus. Il suo approccio alla *lingua ideale*, più che alla tradizione dell'empirismo, fa pensare alla filosofia medievale (*l'Arts Magna* di Raimondo Lullo, terziario francescano ed esoterista ai confini tra follia e genio), al razionalismo (la *Characteristica Univesalis* di Leibniz, che sperava di individuare un alfabeto del pensiero, un insieme di concetti combinabili per calcolare meccanicamente ogni verità), alle lingue artistiche create per gioco e per piacere (una recente è la *Toki Pona*, una lingua artificiale basata sui principi della religione taoista)... insomma, non proprio un canone di tipo analitico!

Il *secondo* Wittgenstein, quello maturo e anzi morente delle *Ricerche* (2014), cambia totalmente strada, crea la *filosofia del linguaggio ordinario*: abbandonata l'idea della necessità di una *lingua ideale*, Wittgenstein sottolinea l'importanza del *senso comune*, un principio caro a Moore (1925), altro padre della filosofia analitica, e tipico nella filosofia scozzese da Reid (1996) in avanti, tanto che pure oggi il rispetto per il senso comune è imprescindibile nella filosofia analitica (sono inaccettabili non solo le teorie filosofiche che vanno contro i dati delle scienze, ma anche quelle troppo lontane dalle comuni intuizioni sulla natura del mondo). Wittgenstein sostiene che i problemi filosofici emergano sì da fraintendimenti del linguaggio, ma non perché sia necessaria una lingua artificiale non ambigua, al contrario, sono i filosofi che creano i loro problemi dimenticando il vero significato che le parole hanno nell'uso quotidiano: non esistono problemi filosofici, la risposta ai problemi l'hanno sempre saputa, ora devono riscoprirli analizzando attentamente l'uso delle parole.

I principi della filosofia del linguaggio ordinario suonano naturali all'interno della filosofia analitica. Sono una estensione del principio del contesto di Frege: il significato delle parole non è comprensibile in sé, lo è nel contesto di una frase, e Wittgenstein aggiunge di una frase *del linguaggio quotidiano*, non qualche altro contesto artificiale; l'errore dei filosofi sta nel prendere le parole in isolamento, in modo astratto, o nell'usarle in contesti problematici. Mentre i filosofi si chiedono «cosa vuol dire natura?» e cercano invece di rispondere alla domanda «cos'è la natura?»

Wittgenstein fa il contrario, si chiede «quando uso questa parola? In che condizioni? Quando direi l'opposto? Quando sarei insensato nel dirlo? Come reagirei sentendolo?», perché la natura è ciò che si dà quando io uso quella parola: indagare la natura vuol dire indagare il significato della parola «natura» e indagare il significato vuol dire indagare il modo in cui uso la parola.

Ciò nonostante, il rapporto tra filosofia del linguaggio ordinario e filosofia del linguaggio ideale è stato ed è problematico. Non devo farne una caricatura, come se la filosofia analitica fosse davvero una scuola coerente in cui tutti sottoscrivono ad un paio di principi indiscutibili. In realtà, non paiono esserci confini netti tra la filosofia analitica e la non-analitica.

Penso a Paul Thagard, uno dei più raffinati filosofi della scienza viventi. Thagard propone un tipo di *naturalismo* che rifiuta l'analisi del linguaggio tramite la logica e il rispetto per il senso comune; favorisce invece una filosofia dedicata alla teorizzazione basata sui dati delle scienze e alla revisione dei concetti popolari che non trovano riscontro. Come Chomsky, Thagard rovescia l'antipsicologismo di Frege, nega che il linguaggio sia un filtro necessario allo studio della mente, al contrario è la mente, la psicologia, che guida lo studio del linguaggio, non è la filosofia della mente che dipende dalla filosofia del linguaggio, è l'opposto (Thagard 2012, 2019). Questo approccio è radicale nel negare i presupposti individuati; eppure, è ancora filosofia analitica, è una corrente interna oggi popolare.

Cosa posso dire del pragmatismo? I più grandi filosofi analitici dell'epoca successiva a Wittgenstein sono Quine e Putnam. E non sono analitici, sono *neo-pragmatici*. Putnam non si è fatto problemi a scrivere su Habermas, il più importante e popolare filosofo continentale vivente; e Habermas, di contro, pur agendo all'interno della teoria critica, sostiene una filosofia politica che può essere considerata analitica e una epistemologia che pare pragmatica (Habermas 2001, 2013). Brandom, uno dei maggiori analitici viventi, ma anche lui neo-pragmatico, ha scritto un libro su Hegel (Brandom 2019). Hubert Dreyfus, uno dei pochi filosofi ad aver lavorato al MIT (la più importante università del mondo e tempio indiscusso della tecnologia), ha addirittura basato la sua filosofia su Heidegger (Dreyfus, dal punto di vista degli analitici, dovrebbe essere l'equivalente di un satanista) (Dreyfus & Dreyfus 1986).

E non esiste solo un pragmatismo analitico, ma anche una fenomenologia analitica, una ermeneutica analitica, un marxismo analitico... E la metafisica e la filosofia della religione, un tempo rifiutate dal positivismo logico come insensatezze, hanno oggi un ruolo di primo piano nella filosofia analitica, tanto che non tutti lavorano più sempre e solo in continuità con la scienza (alcuni, come il logico e maggior filosofo protestante vivente Plantinga, hanno proposto estreme revisioni della pratica scientifica) (Plantinga 1991).

Se la tradizione analitica sta iniziando ad accettare filosofi, temi e metodi della filosofia continentale, quest'ultima si sta muovendo in direzione opposta. Kant, sostenendo la conoscibilità del mondo solo tramite gli schemi mentali, ha giustificato una frequente critica analitica: l'idea che i filosofi continentali siano incapaci di parlare di due oggetti qualsiasi senza che ci sia in mezzo anche un umano che li osserva e commenta. I filosofi del *realismo speculativo*, pur nell'enorme diversità di vedute (è un gruppo di quattro filosofi, sono stati capaci di creare sette sottoscuole diverse), hanno preso molto sul serio questa critica, anzi alla lettera: rifiutano Kant, riprendono Hume, interpretano Heidegger rifiutando lo statuto speciale del soggetto rispetto all'oggetto, presentano le interazioni uomo-oggetto come per nulla eccezionali rispetto alle altre interazioni oggetto-oggetto, rendono l'uomo un oggetto tra gli oggetti (Bryant et al. 2011). Si rifiuta il relativismo continentale per ritornare ad una visione del mondo come indipendente dai costrutti mentali, un tema cavalcato anche da filosofi italiani contemporanei come Maurizio Ferraris (2014),

creatore di un *nuovo realismo* che è l'effetto del realismo speculativo sull'ermeneutica e che è stato apprezzato anche in ambito analitico.

Il colpo di grazia: Quine e Kripke

Quine non è solo il più grande esponente della filosofia analitica del dopoguerra, è anche quello che ha rischiato seriamente di farla finire, l'organizzatore del più pericoloso attentato alle basi concettuali della filosofia analitica così com'era all'epoca: colui che ha messo in dubbio l'esistenza stessa delle proposizioni analitiche e la possibilità di comprendere il pensiero tramite il linguaggio. È un argomento sul quale voglio soffermarmi a lungo perché lo riprenderò più volte.

Inizio con il problema della distinzione tra proposizioni analitiche e sintetiche, che Quine discute nel suo paper più importante, *Two Dogmas of Empiricism*.

Quand'è che una proposizione risulta essere analitica? Una proposizione è analitica in virtù del suo significato, lo è quando le sue componenti hanno lo stesso significato come in «quello scapolo è un uomo non-sposato». Ma come faccio a capire quando due parole hanno lo stesso significato, quando sono sinonimi? Non posso dire che sono sinonimi le parole che compongono una proposizione analitica perché sarebbe circolare. Non posso dire neanche che due parole sono sinonimi se sono sostituibili in tutti i contesti rilevanti: «animali con l'omaso» e «animali con l'abomaso» sono scambiabili ma solo in virtù del fatto che hanno identici riferimenti (entrambi indicano tutti e soli i ruminanti, ciò che dico di uno lo posso ripetere con l'altro), non in virtù del loro significato, che è diverso. Devo allora aggiungere la nozione di necessità: *scapolo* e *uomo non-sposato* sono sinonimi solo se tutti e soli gli scapoli sono necessariamente uomini non-sposati. Ma allora: quand'è che una proposizione è necessaria? Non posso dire né che lo è quando è analitica, né che lo è quando le sue componenti sono sinonimi, perché sono entrambe circolari. C'è una sola alternativa: devo appellarmi ai fatti del mondo, al modo in cui termini sono usati, per esempio posso affermare che sono analitiche le proposizioni che sono chiamate tali dagli esperti, o che sono analitiche quelle proposizioni che i parlanti non sono disposti a modificare o correggere in quanto hanno un ruolo centrale nelle loro credenze. Ma se dico così non sto più parlando di proposizioni analitiche, perché risulterebbero tali in virtù di fatti esterni, non del loro significato, lo sarebbero al modo delle proposizioni sintetiche. Quine conclude così che la verità dipende sempre sia da fattori linguistici che extralinguistici, e quindi che non c'è neanche una vera distinzione tra proposizioni analitiche e sintetiche, ogni proposizione ha sia elementi di analiticità che di sinteticità. (Quine 1951)

Ma questa è solo la punta dell'iceberg: nel suo capolavoro *Parola e Oggetto*, scritto pochi anni dopo i *Two Dogmas*, Quine prende una posizione ancora più radicale.

Quine immagina che un linguista anglofono vada a vivere con una tribù che parla lo sconosciuto linguaggio *arunta* per impararlo e tradurlo. Ecco un gruppo di cacciatori che vedono un coniglio e lo indicano urlando «gavagai!». Ma: cosa vuol dire *gavagai*?

La traduzione più immediata è «coniglio»: il suo riferimento, l'oggetto nel mondo verso cui punta, è chiaramente il coniglio, quindi pare essere quello il suo significato concettuale. Epperò, se i membri della tribù avessero invece inteso dire «cibo» o «stato temporale di un coniglio» o «parti non-distaccate di coniglio» o «esempio di coniglio» il linguista avrebbe assistito agli stessi identici comportamenti. *Gavagai* potrebbe voler dire addirittura qualcosa come «sta per arrivare una

tempesta» (c'è la superstizione, nella tribù, che l'apparire di un coniglio preannunci maltempo), comunque non lo si può distinguere. Per quanto alcune di queste ipotesi di traduzione possano essere scartate in base agli altri usi di *gavagai*, comunque continuerebbero sempre ad esistere infinite possibili concettualizzazioni del termine: se il suo uso è causato dal concetto, e io non posso osservare altro che l'uso, allora, essendoci infiniti concetti che possono causare quello stesso medesimo uso, non potrò neanche mai individuare quale specifico concetto c'è nella testa di chi mi parla.

Si potrebbe pensare che la confusione possa esserci solo tra parole che indicano lo stesso ente. «Coniglio» e «parti non-distaccate di coniglio» paiono avere una uguale *estensione* (un uguale insieme di riferimenti): tutti gli oggetti indicati da «coniglio» sono anche quelli indicati da «parti non-distaccate di coniglio»; non c'è un oggetto che, essendo un coniglio, non è anche un insieme di sue parti non-distaccate, tanto che si potrebbe usare un'espressione come una definizione dell'altra, sono cioè *logicamente equivalenti*, l'essere uno è condizione *necessaria e sufficiente* ad essere anche l'altro. Ciò nonostante, sarebbe un errore credere che siano espressioni con *identico* significato, che siano solo due modi di dire la stessa cosa: *essere coniglio* non è identico all'*essere un insieme di parti*, sono espressioni che si riferiscono a proprietà diverse, sia pure entrambe possedute dallo stesso oggetto. E comunque questo non lo si può dire per tutti i casi: si può confondere anche «coniglio» e «cibo» che hanno senza dubbio estensioni diverse non solo in quanto a proprietà, oppure si può immaginare una interpretazione contestuale tale che «gavagai» indichi ciò che io intendo con «coniglio» in certi contesti e con «cibo» in altri.

Torno alla tribù, trovo il linguista che, per controllare una precedente ipotesi, va dal capo villaggio e gli chiede «*gavagai* significa *cibo*?» Purtroppo non può funzionare: lui può fare domande al capo villaggio solo tramite la lingua *arunta*, la quale lui può imparare solo ipotizzando il senso dall'uso; anche chiedendo direttamente il significato dei termini non può comunque arrivare ad un'unica interpretazione legittima perché, posta una stessa risposta, può immaginare che l'interlocutore abbia interpretato la domanda in una infinità di modi possibili diversi da quello che intendeva.

Prova con una diversa domanda, «l'apparizione del coniglio è presagio di tempesta?» e il capo villaggio risponde in modo allegorico «no, i conigli sono nostri fratelli.» Come può il linguista capire questo uso astratto del termine *fratello* che è indipendente da qualsiasi dato comportamentale osservabile? Come gestire i casi in cui *gavagai* è usato in modo indipendente dalla presenza fisica del coniglio, per esempio quando si narra un racconto o quando si trovano delle impronte, come si può capire se vuol dire «coniglio» anche in tali contesti? E la sintassi della frase? Com'è possibile capire se *gavagai* sia un nome, un verbo, un articolo, o anche più cose contemporaneamente, visto che la frase può essere divisa in più modi, che posso assegnare alle parole che la compongono diverse funzioni linguistiche, e ottenere infinite strutture logiche possibili ugualmente coerenti?

Quine ha individuato un'indeterminatezza pervasiva nel linguaggio. Non in un linguaggio, o nella lingua naturale, ma in tutti i linguaggi, anche le lingue artificiali, i linguaggi di programmazione, le teorie scientifiche espresse in linguaggio matematico, tutto pare essere legittimamente interpretabile in una infinità di modi, il significato che io assegno ad una certa espressione dipende dalla mia lingua madre e non c'è una interpretazione oggettivamente più corretta delle altre. (Quine 2008)

Sembra che Quine voglia dimostrare l'impossibilità di tradurre un libro. Non è così, vuole evidenziare un problema nel concetto stesso di proposizione. Le frasi «the snow is white» e «la neige est blanche» esprimono la stessa proposizione (esplicitabile come *la neve è bianca*), cioè hanno uno stesso significato indipendente dalla specifica lingua parlata. Quine mette in dubbio la

sensatezza di tale idea: in realtà le frasi non esprimono alcuna proposizione specifica, non c'è alcun fatto in merito; quando il capo villaggio mi dice «i conigli sono nostri fratelli» io non posso sapere se sta dicendo qualcosa di vero o falso perché non posso sapere se la sua frase esprime una proposizione vera o falsa, ci sono infinite proposizioni che potrebbero essere espresse da quella frase.

Ora che Quine e Putnam sono morti il più grande filosofo vivente è Saul Kripke: introverso, impacciato, prodigio della matematica (gli avevano proposto di andare a insegnare ad Harvard, lui ha risposto «la mamma ha detto che prima devo finire le superiori») nonché maestro Jedi, a giudicare dalla sua voce identica a quella di Yoda. Nel libro *Su Regole e Linguaggio Privato* collega Wittgenstein con la teoria della traduzione di Quine in quello che chiama *rule-following paradox*.

Io potrei aver imparato il significato del segno «+» derivando una regola da varie somme: « $2+2=4$ », « $3+7=10$ », « $109+327=436$ ». Immagino ora che esistano due numeri dei quali né io né la mia maestra di matematica abbiamo mai calcolato la somma, come 35 e 26, e che uno scettico venga a dirmi che la regola del segno «+» sia «il risultato di $x+y$ è sempre la somma di x e y , ma se $x=35$ e $y=26$ allora il risultato è 5»: ho io ragione di credere che la mia maestra non volesse effettivamente insegnarmi questa regola? O che il mio io passato non stesse seguendo questa? Pare proprio di no: se il significato comporta una certa regola d'utilizzo che si manifesta in un corso d'azioni, posto che per un certo corso d'azioni esistono infinite regole che possono rendere merito dello stesso corso, allora, posto un certo utilizzo di una parola, mi è legittimo ipotizzare infiniti significati possibili che rendono merito di quell'utilizzo (Kripke 2000).

Mi piacerebbe poter sostenere che il problema sta nel partire da una forma di *comportamentismo*, nel credere che si possa conoscere il significato solo sulla base del comportamento a prescindere dagli stati mentali. Ma non è forse così che ho appreso il significato delle parole da bambino, osservando il comportamento? E comunque, Kripke nega la rilevanza degli stati mentali: anche avessi accesso diretto a tutta la mente, anche potessi analizzare ogni struttura presente nel mio cervello, nulla potrebbe fissare l'interpretazione della parola, un certo suo uso futuro anziché un certo altro, neppure un Dio onnisciente può ricondurre l'uso di una parola ad un'unica regola oggettiva. Quine e Kripke dimostrano che il significato della frase non è riducibile a null'altro, né l'uso né gli stati mentali né qualsiasi altro fatto possono essere tali da collegare una frase ad una univoca proposizione.

Queste argomentazioni sono paurose, sembrano dire che non esiste proprio alcun significato, che il linguaggio non esprima alcuna proposizione, né analitica né sintetica, e se è così ogni speranza di poter comprendere la mente tramite l'analisi del linguaggio va cestinata.

Il lavoro di Quine ha lasciato segni indelebili nella filosofia analitica. Mentre i positivisti logici, rifiutando Kant e seguendo Hume, credevano che ogni proposizione analitica fosse conosciuta a priori e necessaria e ogni proposizione sintetica fosse a posteriori e contingente, oggi si accetta l'esistenza di proposizioni sintetiche a priori e, tramite il lavoro di Kripke (1999), anche di verità necessarie a posteriori (il che danneggia l'argomentazione di Quine, visto che partiva dal presupposto che ogni verità necessaria fosse analitica). Quine inoltre ha contribuito a creare una teoria *olistica* del significato che costituisce una estrema generalizzazione del principio del contesto di Frege: non solo una parola è comprensibile solo all'interno di una frase, ma quella frase è comprensibile solo all'interno di un intero linguaggio (Quine 1951).

Oggi, nonostante la difficoltà del distinguere la necessità dalla contingenza e la conoscenza a priori dalla conoscenza a posteriori, la maggior parte dei filosofi accetta l'esistenza di una distinzione tra

proposizioni analitiche e proposizioni sintetiche (Bourget & Chalmers 2014). Quine dimostra che è scorretto definire l'analiticità in base alla sinonimità, e la sinonimità in base alla intersostituibilità dei termini, ma queste non sono le uniche definizioni possibili. Searle, uno dei grandi filosofi del linguaggio viventi, nota che la correttezza di una definizione di *analiticità* verrebbe controllata confrontando la sua estensione con l'estensione degli enunciati analitici: questo non vuol dire automaticamente che una definizione sia possibile, perché l'analiticità potrebbe essere inanalizzabile, ma l'inalizzabilità non implica l'inesistenza di proposizioni analitiche, anzi, mostra che si conosce l'estensione degli enunciati analitici anche senza averne una definizione (Searle 2009).

E riguardo l'indeterminatezza, il fatto che sia impossibile trovare una relazione *uno-a-uno* tra frasi e proposizioni espresse, né Quine né il *Kripkenstein* (il Wittgenstein così com'è recepito da Kripke) lo credono dovuto ad errori di qualche tipo, anzi accettano i paradossi, ma loro stessi li superano in modo pragmatico: per Quine dire che «gavagai» significa *coniglio* non è una traduzione ma una questione di buon senso, è accettabile finché la proposta di traduzione è coerente con il comportamento; per il Kripkenstein similmente bisogna accettare che una certa parola abbia un certo significato finché non abbiamo motivo di dubitarne, cioè finché qualcuno non reagisce all'uso della parola in modo inaspettato. Kripke non mette in dubbio che io, nel momento in cui parlo, so quale proposizione sto cercando di esprimere, so a che cosa sto cercando di riferirmi.

Ma le ambizioni di Frege sono oggi abbandonate, nessuno crede di poter risolvere i più complessi problemi filosofici a suon di analisi e riduzione ad enunciati analitici: l'analisi può chiarire i concetti, ma per chiarire il ritratto del mondo serve qualcosa di più. Va bene: cosa? Come nota Dummett (2001), se il linguaggio è lo specchio del pensiero, allora io posso correggere la mia concezione della realtà agendo sul linguaggio; ma se non lo è, come posso agire sul pensiero in sé?

Mi sembra che queste due visioni, quella di Frege e quella post-Quine, si riferiscano a due processi diversi, a due modi di lavorare della mente. Frege parte dall'idea che il senso della parola determini il riferimento e preceda il pensiero, per questo si può cambiare il ritratto del mondo cambiando il significato delle parole. L'uomo, però, è capace anche dell'atto contrario, può creare concetti partendo da esperienze di oggetti, e se si vuole indagare questo aspetto della mente non si può partire dal linguaggio, il metodo di Husserl in questo caso è più logico di quello di Frege.

Cos'è la filosofia analitica?

Quindi:

- 1- non tutti i filosofi analitici rifiutano Hegel ed Heidegger (e non tutti i continentali li accettano);
- 2- non tutti i filosofi analitici ritengono che il linguaggio sia il filtro fondamentale tramite il quale studiare la mente, anzi sostengono anche forme di psicologismo (sono invece gli strutturalisti continentali che, a volte, hanno sostenuto questa importanza ermeneutica del linguaggio);
- 3- pochi analitici contemporanei perseguono ancora il progetto di Frege, anzi si riconosce che molti concetti filosofici sono inanalizzabili;

4- non tutti gli analitici ritengono che la matematica e il rispetto dell'autorità della scienza siano elementi fondamentali dell'attività filosofica (ci sono invece continentali che lo sostengono, da Leibniz fino ad un contemporaneo come Badiou).

Ma allora cosa vuol dire essere *analitici*? Be', cosa vuol dire *gavagai*?

A volte, quando sono in bagno, all'apice dello sforzo, mi diverte esplorare le genealogie accademiche: Leibniz è stato il maestro di Jacob Bernoulli, che lo è stato di suo fratello Johann, che fu maestro di Eulero, che fu maestro di Lagrange, che lo fu di Fourier, che lo fu di Dirichlet, che lo fu del finitista Kronecker, che fu maestro di Hensel, che lo fu di Hasse, che lo fu di Wolfgang Franz, che lo fu del premio Nobel per l'economia Reinhard Selten, che lo fu di Eric Van Damme che è ancora vivo e vegeto. Io penso che la filosofia analitica sia qualcosa del genere: è una genealogia spirituale. L'essere analitici (come continentali o qualsiasi altra cosa) non consiste tanto nello sposare certe posizioni precise, quanto nel sentirsi parte di un mondo figlio di un certo modo di pensare, e nel riconoscere i propri padri di conseguenza.

La filosofia analitica è uno stile basato sulla chiarezza e la precisione logica, sì, ma lo stile non esiste indipendentemente da coloro che lo hanno utilizzato. Io sento che gli uomini del passato, e certi uomini in particolare, hanno lasciato a tutti coloro che hanno avuto la fortuna di nascere dopo di loro come un potere, un artefatto antico, con il desiderio che questo possa giovare all'umanità, che possa dare speranza e saggezza. Ognuno si guarda indietro e vede uomini diversi, ma è l'autorità riconosciuta, o meglio, il senso di responsabilità verso questi, che crea la tradizione.

LETTURE PER APPROFONDIRE

Chi vuole studiare filosofia per conto proprio può iniziare a orientarsi tramite Wikipedia: si cerca «Western Philosophy» e da lì si inizia a leggere, di collegamento in collegamento, le pagine delle varie scuole, dei vari settori, delle principali teorie, e dei pensatori più importanti.

Il passo successivo è visitare la Internet Encyclopedia of Philosophy e la Stanford Encyclopedia of Philosophy: la prima molto semplice, la seconda più complessa ma anche più rigorosa e aggiornata, sono enciclopedie gratuite curate da professionisti del settore; gli articoli della SEP, in particolare, sono sufficienti ad avere un'ottima infarinatura di base sugli argomenti scelti.

Per assistere a corsi o lezioni di filosofia, servizi come HarvardX, Harvard Online Learning, MIT OpenCourseWare, Stanford Online, Open Yale Courses permettono di seguire gratuitamente lezioni tenute da professori delle più importanti università del mondo, e alcuni rilasciano (a pagamento) certificati di partecipazione. Ovviamente molte università caricano le proprie lezioni anche su YouTube, mentre gli studenti diffondono i propri appunti online.

Si continui quindi leggendo le ricerche scientifiche sugli argomenti di proprio interesse. Anzitutto, bisogna trovare una bibliografia: Philpapers raccoglie bibliografie per 5671 diversi argomenti, e ognuno di questi può richiedere mesi o anni per essere compreso a pieno (chiaro quindi che nessun uomo può avere dominio di tutta la filosofia). Se si trova un ricercatore di proprio interesse, è possibile rimanere aggiornati sulle sue ricerche seguendolo su Philpeople o su altri social come Researchgate e Academia.edu. Queste ricerche saranno spesso scaricabili gratuitamente da questi siti oppure da Philarchive.

Se non si riesce a trovare una specifica ricerca, si può continuare a cercarla tramite motori di ricerca accademici come Google Scholar, BASE, CORE, JURN, Microsoft Academic Research, Zetoc, anche se il mio preferito è CiteSeerX.

E se anche così non si trova niente, o se la si trova solo a prezzi elevati, l'esistenza di servizi quali Library Genesis e Sci-Hub è un segreto di pulcinella tra i ricercatori indipendenti e non, chi ritiene etico utilizzarli troverà facilmente online come fare.

STORIA DELLA FILOSOFIA

Non ho parlato fin qui di filosofia antica, ma vale la pena accennarvi visto che molti continentali, rifiutando la tradizione di Platone e Aristotele, vanno a pescare i filosofi precedenti o appena successivi, cioè i presocratici e gli ellenici, o anche i filosofi non-occidentali. La più famosa raccolta di frammenti presocratici è Diels & Kranz 1903. Si vedano poi Gorgia 2010, Sesto Empirico 1999, Pirrone 1981, Epicuro 2007, Panezio 1997, Posidonio 2004. Per i non-occidentali, le basi assolute sono Averroè 1994, Maimonide 2013, Confucio 2006, Lao-Tzu 2013, Shankaracharya 2015.

Per il tomismo, la filosofia medievale rimane fondamentale. L'opera più importante è d'Aquino 1995-2013, enorme e spesso anche molto costosa (l'edizione citata è lunga più di 16.000 pagine, e costa più di 1.500 euro); più abbordabile invece d'Aquino 2016. Il secondo pensatore più importante per la cristianità medievale è Agostino, con la trinità di opere 2012, 2013, 2015. Quindi, tra i vari altri Padri, Dottori, Scrittori Ecclesiastici, si vedano: Origene 2012, Proclo 1994, d'Aosta 2002, Eriugena 2013, Scoto 2008, di Occam 2015, Alberto Magno 2007, Boezio 2014. Estendendo ai rinascimentali, importanti per la cristianità postmoderna, si vedano Erasmo 2005, More 2016, Cusano 2017, Ficino 2013, Pomponazzi 2013, de Montaigne 2014, della Mirandola 2013, Campanella 1997, Bruno 1985. Infine, per i pensatori recenti di questa tradizione: Garrigou-Lagrange 2012 e 2015, Fabro 2005, 2006 e 2010, Gilson 1984, Rahner 1992, Lonergan 2001.

Passando alla filosofia moderna, il manifesto del razionalismo è il brevissimo Cartesio 1637. Lo stile razionalista è perfettamente realizzato in Spinoza 1677: partendo da assiomi, Spinoza deriva per teoremi la sua etica, dimostrandola al modo della geometria, anche se i suoi assiomi non sono autoevidenti come quelli matematici, né le sue dimostrazioni altrettanto rigorose. Wolff 2011 prosegue la tradizione razionalista.

La bibliografia di Leibniz è sterminata, e molte delle sue opere rimangono inedite: in 1666 nota la vicinanza concettuale di necessità e obbligo, possibilità e permissibilità, un'intuizione che anticipa di tre secoli la logica modale; in 1720 espone in appena una decina di pagine la sua metafisica, un testo che per la sua brevità è spesso incluso come appendice in 1710, dove discute invece il problema del male; infine, in 1765 commenta frase per frase Locke 1690. Il miglior libro su Leibniz è Antognazza 2015, davvero un'opera che richiede tutta una vita. E per curiosità, Small Monster Studio ha sviluppato un videogioco, chiamato appunto Leibniz, dedicato al sistema binario e all'arte barocca.

I due testi più importanti dell'empirismo sono Hume 1739 e 1748: il primo molto lungo, il secondo molto breve, sono il cuore di tutto il canone della filosofia analitica. Molto diverso, ma sempre empirista, è Berkeley 1710, che nega l'esistenza di un mondo esterno conoscibile.

Le tre grandi opere di Kant sono 1781, 1788, 1790, dedicate rispettivamente all'epistemologia, all'etica e all'estetica: la prima è la più complessa, bisogna leggere lentamente ogni frase chiedendosi cosa può aver portato Kant a dire ciò che ha detto; la seconda, invece, è la più breve, un ottimo punto per iniziare a conoscere l'idealismo trascendentale.

Tra gli idealisti, consiglio Fichte 1992 e Schelling 1990, due testi minori ma piacevoli e molto semplici rispetto alle loro altre opere. Poi si potrà passare a Hegel 1807, un libro scritto in modo alieno e che richiede enorme pazienza. L'influenza di Hegel è ben visibile in Italia (Gentile 1916, Croce 1909, Spaventa 2009), in Germania (Feuerbach 1841), in America (Emerson 2005 e 2009), in Inghilterra (Collingwood 1971) e, a dispetto delle apparenze, anche in positivisti come Comte

(2017), in neo-kantiani come Cassirer (2018), e addirittura in scienziati come il fisico e medico Hermann von Helmholtz (2003). Tra i contemporanei, il più spiccatamente hegeliano è Žižek (2012, 2016, 2017).

FILOSOFIA CONTINENTALE E ANALITICA

Il libro centrale di tutto il canone continentale è Heidegger 1927, opera di lettura difficilissima, serve molta forza di volontà per arrivare in fondo.

L'opera più importante per quel che concerne la fenomenologia è Husserl 1900. Tra i suoi seguaci si vedano: Levinas et al. 2008, Ricoeur 2016, Merleau-Ponty 2003, Jaspers 2006, Marion 2018. Per l'associata etica personalista si vedano invece Scheler 2013, Mounier 2004, Schütz 2018, Henry 1997, e soprattutto von Hildebrand 2003.

Gli esistenzialisti si esprimono spesso tramite forme narrative, e per questo sono piacevoli da leggere: Sartre 1938, Camus 1942, Kierkegaard 1843, Dostoevskij 1866, Nietzsche 1883. Per l'esistenzialismo femminista, si vedano de Beauvoir 1949, Haraway 2018, Butler 2014 e 2017. Per l'esistenzialismo religioso invece, Buber 2011, Tillich 1996-2006, Sestov 2005 e 2016, e soprattutto Berdjaev 2017.

I testi fondativi dell'ermeneutica sono Schleiermacher 1985 e 2000; il più importante tra i moderni invece è Gadamer 1960. Altri moderni che hanno declinato l'ermeneutica in modo personale sono Dilthey 2007, Bloch 2019, Jameson 2007.

Il testo più importante della tradizione marxista è ovviamente Marx 1867; essendo però un libro lungo e noioso, è forse meglio iniziare con Marx 2018. Tra i seguaci di Marx si vedano Gramsci 2014 e Althusser et al 2006.; quest'ultimo, Althusser, oltre ad essere il maggior marxista del secolo scorso, è anche tristemente noto come uxoricida. In merito al discutibile effetto del marxismo sulla critica letteraria, si vedano invece Lukács 1997 e 2004, e Barthes 1981, 2002 e 2003.

Passando a strutturalismo e post-strutturalismo, della sterminata bibliografia di Foucault credo che i testi più interessanti siano 1998, 2014 e 2013-2019, dedicati rispettivamente alle cliniche come simbolo di potere sulla vita, alle carceri, e alla sessualità. Tra i vari altri pensatori postmoderni si vedano, per gli aspetti generali Deleuze & Guattari 2002 e Marcuse 2001, per gli aspetti semiologici Baudrillard 2008 e Jakobson 1986, per quelli antropologici Lévi-Strauss 2008. I due pensatori più difficili da leggere di questa scuola sono senza dubbio Derrida (1967 e 2002) e Lacan (2002), veri odiatori della chiarezza. Il padre di questa scuola di pensiero, a mio parere, è Bachelard, autore di opere originalissime come 2006, 2015, 2016, 2018 e 2020.

Per l'ambito della teoria critica, si vedano Horkheimer 2002 e 2003 e Adorno 2015; per l'estetica, Benjamin 2011 è uno dei testi più importanti (e più brevi) di questo settore filosofico; per l'aspetto psicologico, invece, Fromm 2018. Poiché la teoria critica è essenzialmente una teoria sociologica, può essere buona cosa contestualizzarla con l'aiuto di manuali di sociologia quali Berger & Berger 1997, Mitchell 1971, o Collins 1980.

L'opera che ha fondato la moderna filosofia del processo è Whitehead 1929 (anche se già prima vi erano filosofi che sostenevano principi simili, come Bergson 1907); essendo però un libro complesso sia da leggere che da reperire, ci si può accontentare anche di Whitehead 2015. Tra i seguaci di Whitehead sono da segnalare psichiatri anti-psichiatria quali Laing (1971) e Szasz (2003).

L'oggettivismo, come l'esistenzialismo, si esprime spesso in forma narrativa, ma purtroppo Ayn Rand, con opere come 2004, si dimostra meno talentuosa di Sartre. Fortunatamente, esiste una versione cinematografica, Vidor 1949. C'è un certo numero di seguaci di Rand, come Leonard Peikoff (che paragona la sua maestra direttamente a Platone e Aristotele), ma sono scarsamente influenti al di fuori della politica di destra americana.

Peirce 2011 è il libro che contiene tutto ciò che si può desiderare per quel che riguarda il pragmatismo, e se non bastasse lo si può affiancare a James 2009 e Dewey 2014. Tra i pragmatisti recenti, si vedano almeno Sellars 2007 e 2013 e Rorty 2004. Da recuperare anche la tradizione del pragmatismo italiano, che vantava esponenti quali Papini (2018), Vailati (1980) e Calderoni (2007).

Infine, per la rinascita dei movimenti realisti nella filosofia continentale basti leggere Ferraris 2014 (altri, come Harman e Meillassoux, non hanno avuto altrettanto successo a livello internazionale); si noti però che queste tendenze non si distaccano più di tanto dal ben più vecchio Holt et al. 1912.

Per la filosofia analitica, i già citati Quine 2008 e Kripke 1999 e 2000 sono tre libri fondamentali. Per un'introduzione alle idee e alla vita di uno dei padri della filosofia analitica, Russell, sarà facile trovare sulle bancherelle la raccolta Paci 1972, la quale rivela anche un Russell scrittore di racconti tutt'altro che rigido e austero.

Di Wittgenstein, è necessario leggere tutto. Si può iniziare con Wittgenstein 2000, appunti dettati agli studenti, più chiari e semplici delle altre opere. Ai due grandi libri Wittgenstein 2009 e 2014 bisogna far seguire 1999, che può essere considerato una sorta di terza fase del suo pensiero; questi libri sono strutturati in modo molto chiaro, ma la loro scrittura non lo è affatto, ricorda Nietzsche. Per il pensiero etico ed estetico si veda invece Wittgenstein 1967. Per conoscere l'appassionante vita di questo ribelle, io ho adorato Wittgenstein 2005 e l'epistolario 2021. Purtroppo, non ci sono registrazioni audio o video di Wittgenstein in vita (uno dei più grandi compositori viventi, Steve Reich, che lavora campionando registrazioni, ha cercato a lungo ma non è riuscito a trovare alcunché); ciò nonostante, le sue foto abbondano: ecco il meraviglioso Nedo 2018. Esiste anche un film, Jarman 1993, dedicato a Wittgenstein.

Poiché la filosofia analitica è il tentativo di comprendere la mente tramite un'analisi matematica del linguaggio, per poter davvero capire a fondo Frege 2007, 2019, 2021 può essere utile approfondire la linguistica e la logica formale. Per la linguistica, il manuale sul quale ho sudato sangue è Berruto & Cerruti 2011. Per gli aspetti più propriamente filosofici, invece, si veda Marconi 2008. Per la logica, consiglio il libro di un ex-studente della mia università, Berto 2008, dal quale si può poi passare a Lemmon 1986.

Introduzione ai ragionamenti di epistemologia

Analisi del concetto di conoscenza

Inizio con l'epistemologia come Cartesio, ma non per questo intendo dire che l'epistemologia sia la filosofia prima. Per Lévinas (e forse Platone) la filosofia prima è l'etica (Lévinas 2016), per Peirce l'estetica, per Russell la logica, e tutti hanno le loro ragioni. Invero, non c'è una filosofia prima, ogni discorso filosofico contiene al suo interno elementi etici, metafisici, linguistici, che andrebbero indagati tutti insieme. Il motivo per cui ho scelto di iniziare con l'epistemologia è che mi permette di affrontare delle teorie che sono fondamentali per le argomentazioni che esporrò negli altri settori.

E poiché seguo la prospettiva analitica, devo iniziare con l'analisi del concetto di *conoscenza*. Non richiederà grande sforzo visto che una tradizione che risale fino a Platone tramanda un'analisi già eccellente: *credenza vera giustificata* (Platone 2009, nel Teeteto e nel Menone).

Che queste tre dimensioni siano componenti del concetto di conoscenza lo si può intuire con facili esempi: una conoscenza è una credenza perché sarebbe assurdo dire «io so che piove, ma non lo credo» (cfr. Moore 1993); non è una credenza qualsiasi, ma deve essere una credenza vera, perché io posso dire «credevo piovesse, ma non era vero», ma non posso dire «sapevo che stava piovendo, ma non era vero»; ed è infine una credenza giustificata, cioè devo avere un buon motivo per crederla vera, visto che se io decido di credere che piove perché ho tirato una moneta ed è uscita testa, ed è vero, comunque non è una conoscenza.

Si tratta di un'analisi convincente, implicita in tutta la filosofia occidentale. Si possono fare ulteriori test, si può provare a sostituire «credenza vera giustificata» in varie frasi in cui compare la parola «conoscenza», ma ora non serve: lo scopo non è produrre un sinonimo perfetto e completo (forse non ne esistono neanche), è sufficiente riuscire ad individuare delle componenti del concetto, e così si è fatto. Anche se non è un'analisi perfetta è, rispetto all'analisi ideale, ciò che la fisica newtoniana è per la fisica attuale: non falsa, solo una variante *ristretta*, una descrizione vera per molti casi rilevanti anche se non per tutti.

Da questa analisi discendono i due settori dell'epistemologia: il primo è la *teoria della verità*, il secondo la *teoria della giustificazione*. La teoria della verità è ancora analisi, è il tentativo di analizzare e definire la parola «verità»; la teoria della giustificazione invece è l'anima dell'epistemologia e indaga quali sono le condizioni alle quali è legittimo ritenere una certa credenza come vera (a prescindere dal suo esserlo davvero, riguarda la conoscenza e non la natura).

Epistemologia: teoria della verità

Primo ragionamento: una difesa del realismo

Il ragionamento (assolutamente non rigoroso) che segue fa da schema al capitolo: i vari paragrafi sono argomentazioni, spiegazioni, difese delle premesse. Si tratta in questo caso di un ragionamento probabilistico, non cogente: mi limito a soppesare le argomentazioni proposte a favore e contro le varie tesi.

- P1** ci sono prove contro l'antirealismo;
- P2** non ci sono prove a favore dell'antirealismo;
- P3** il pluralismo non è un'alternativa al realismo;
- P4** il deflazionismo non è un'alternativa al realismo;
- C** quindi il realismo è la teoria più probabile.

La tesi realista: il corrispondentismo

Nel vangelo di Giovanni, Gesù dice a Pilato «Per questo io sono nato e per questo sono venuto nel mondo: per dare testimonianza alla verità. Chiunque è dalla verità, ascolta la mia voce», Pilato risponde «Cos'è la verità?» e se ne va senza attendere risposta (Gv 18, 37-38). Un grande filosofo del linguaggio, Austin, interpretando la domanda come uno sberleffo, ha commentato che Pilato era avanti rispetto al suo tempo perché è un errore ridurre la verità ad una qualche altra categoria (una sostanza, una qualità, una relazione), bisogna invece riflettere sugli usi della parola (Austin 1950).

Se c'è una cosa che non sopporto (a parte i biscotti rotti, la carta igienica ruvida, gli inquilini del piano superiore che fanno ginnastica a mezzanotte, e un'altra metà degli eventi dell'universo) è iniziare la riflessione su una certa parola con la definizione del dizionario: è un modo pigro e non informativo di procedere. Ma d'altro canto, cos'è un dizionario se non una gran raccolta di possibili analisi, che è proprio quello che sto cercando di scrivere? Ordunque: verità, «ciò che corrisponde esattamente ad una determinata realtà». Perfetto, problema risolto, avanti il prossimo.

Ho letto una cosa più profonda di quanto possa sembrare, questa definizione rappresenta davvero la più comune e influente teoria della verità, che, in barba ad Austin, è proprio una relazione: una affermazione è vera se corrisponde ai *fatti* del mondo, cioè la verità è una relazione tra le frasi e la realtà.

Un'*affermazione* è una frase che riguarda dei fatti, e quindi che può essere o vera o falsa, cioè che esprime una proposizione. Sarebbe assurdo dire che un ordine come «chiudi la finestra!», una domanda come «è aperta la finestra?» e altre frasi di questo tipo siano vere o false. Perché? Perché non riguardano in alcun modo i fatti. Cos'è un fatto? Il mondo è l'insieme di tutti i fatti (Wittgenstein 2009), quindi i fatti sono tutto ciò che è reale. E cos'è la realtà?

Io vedo il tavolo sul quale mi sto appoggiando. Questo tavolo è, anzitutto, nella mia testa: è la mia rappresentazione di un tavolo. Ma non penso che sia *solo* nella mia testa: non è una allucinazione, non è *solo* veduto, è qualcosa che continua ad esistere in modo indipendente da me; io potrei scomparire di colpo dall'universo a causa di una fluttuazione quantistica e questo tavolo continuerebbe a stare qui, anche se nessuno lo interpreterebbe come un tavolo. Questo è un fatto reale: è ciò la cui esistenza è indipendente dalla mente.

Se ci sono cose indipendenti dalla mente, se c'è un mondo al di fuori di me e della mia esperienza, allora c'è una distinzione tra gli oggetti *così come sono* e gli oggetti *così come sono pensati*. Questo lo diceva Kant (1781) e lo dicono tutti i non-filosofi di tutti i tempi. Ma forse non tutti si sono resi conto che è proprio questa divisione che permette di dubitare, di dubitare di tutto, nel modo iperbolico di Cartesio: infatti, se la realtà non fosse indipendente dalla mia mente, allora l'idea di un mondo radicalmente diverso da come lo conosco è insensata; ma se la realtà è indipendente dalla mia conoscenza, allora posso immaginare che il mondo reale sia completamente diverso da tutto ciò che posso conoscere, pensare o immaginare. Se accetto che la realtà è indipendente dalla mia mente, allora devo anche accettare la *fallibilità* della mia conoscenza, devo accettare che il fatto che io sia in grado anche solo di immaginare le cose così come sono non è affatto una necessità logica, il mondo potrebbe essermi incomprensibile e inimmaginabile.

Questa posizione è chiamata *realismo metafisico*; la sua negazione, cioè l'idea che non vi sia un mondo indipendente dalla conoscenza d'esso, è detta *antirealismo*. Ci sono tipi di realismo che possono non fare coppia con il corrispondentismo (e.g. Morris 2005), ma si tratta di un matrimonio facile e molto ben riuscito, chi rifiuta uno tende a rifiutare anche l'altro.

La tesi antirealista 1: teorie epistemiche

Nel corrispondentismo la verità resta fuori di me, una affermazione su un oggetto è vera se l'oggetto è proprio così come dico. Tutte le teorie che negano ciò sono dette *epistemiche*, cioè *basate sulla conoscenza* in quanto fanno collassare verità e conoscenza, la realtà così com'è e la realtà come posso conoscerla.

Ho detto che i criteri della conoscenza riguardano la giustificazione, mi dicono quando posso affermare di conoscere qualcosa senza però assicurarmi che la tal cosa sia *fattualmente* vera. Nelle teorie epistemiche non è così, i criteri della verità sono quelli della giustificazione, il fatto che io possa dire di conoscere qualcosa è sufficiente non solo a riconoscere la tal affermazione come vera, ma a renderla vera.

Per la teoria corrispondentista l'affermazione «sta piovendo» è come un dito, un dito puntato fuori dalla finestra attraverso la quale controllare se sta davvero piovendo oppure no. Ma se quella finestra non c'è? Io posso sentire il tacchettio della pioggia, il particolare fruscio delle macchine sull'asfalto bagnato, posso dire che questo è il meglio che ho per determinare se sta piovendo, se è sufficiente posso anche affermare «io so che piove». Un corrispondentista direbbe che questa è

vera conoscenza solo se davvero fuori sta piovendo, anche se senza finestra non posso controllarlo. Ma il teorico dell'epistemicismo non è d'accordo: che senso ha che io ipotizzi un mondo oltre la mia stanza dove c'è qualcosa di oscuro, la pioggia, a cui la mia frase corrisponde? Tale mondo può esserci o non esserci, in ogni caso mi è inaccessibile, non posso riferirmi ad esso neanche volendo; per me la pioggia è ciò che io sento e verifico, sono i rumori che sento, non c'è null'altro che possa intendere quando dico «so che piove», non ci sono altri fatti, che non mi sono neppure concepibili, a cui la proposizione dovrebbe corrispondere.

Poiché la verità, nelle teorie epistemiche, non è fattuale, non è neanche indipendente dalla mente: il concetto epistemico di verità è quello di una verità costruita, una verità che può cambiare man mano che la riflessione procede, un teorema matematico non è vero finché un matematico non costruisce la verità con la dimostrazione di quel teorema.

Se, dopo aver affermato che le affermazioni non sono rese vere dai fatti, faccio il passo ulteriore e aggiungo che questi fatti non ci sono, o che sono irraggiungibili e inimmaginabili, allora divento un antirealista.

L'antirealismo ha un evidente vantaggio: qualsiasi pretesa scettica, qualsiasi tentativo di mettere in dubbio la mia capacità di conoscere il mondo, perde senso perché non c'è più alcun elemento che possa ingannarmi o possa costituire una differenza tra la realtà e le mie credenze giustificate. E non è neanche una idea controintuitiva, chiunque è antirealista nei confronti di qualche parte del discorso: per esempio io sono antirealista riguardo all'esistenza degli unicorni, quando dico che l'unicorno «ha la forma di un cavallo, con un corno sulla fronte» non ho certo la pretesa di conformarmi ad alcun fatto riguardo la realtà, io penso che «l'unicorno ha un corno sulla fronte» sia vero ma non certo perché esistano unicorni nel mondo.

Descrivere l'epistemicismo e l'antirealismo in questi termini li fa sembrare totalmente antiscientifici. Non è così: tra coloro che hanno interpretato la verità come costruita dalla attività umana, o che hanno contribuito a creare gli strumenti poi usati per interpretarla così, ci sono: Norbert Wiener, prodigio della matematica (si è laureato a 15 anni, ha preso il dottorato a 19) e fondatore della cibernetica (la teoria generale del controllo e della comunicazione) (Wiener et al. 1943); Jean Piaget, uno dei più importanti psicologi di sempre e in assoluto il più importante nella psicologia dello sviluppo (Piaget 2016); altri vari psicologi importanti come Watzlawick (2018), von Glaserfeld (2016) e George Kelly (2004); il premio Nobel per l'economia Herbert Simon (1988); il fisico e polimata Heinz von Foerster (Segal 1986), che chiedeva «dov'è la realtà? Puoi farmela vedere?» Cosa accomuna personaggi così diversi? Direi il loro focalizzarsi sull'attività interpretante: mentre nel corrispondentismo una relazione a due tra una proposizione e un fatto è sufficiente a definire la proposizione come vera, in una teoria epistemica serve anche un terzo elemento, colui che interpreta e fa conoscenza, senza che vi sia anche costui non c'è verità, può esserci come no un mondo esterno oggettivo ma non c'è verità.

La tesi antirealista 2: il coerentismo

Le due maggiori alternative al corrispondentismo sono il *coerentismo* e la *teoria della verità pragmatica*.

Voglio esprimere la teoria coerentista con un famoso aforisma: la verità non è una proprietà di singole proposizioni, ma di *sistemi* di proposizioni.

È vero, l'essere umano non riesce mai ad essere totalmente coerente con le sue credenze, si contraddice di continuo; ma è vero anche che non può conoscere proposizioni isolate, conosce insiemi di proposizioni. Ora affermo: «in questa stanza la temperatura è di 25 gradi centigradi» (anche d'inverno non tengo mai di meno). Come lo so? Se posso dire di saperlo, posso dirlo solo in relazione ad una certa altra serie di credenze: che il termometro funzioni in un certo modo, che sia composto da certi materiali con determinate proprietà fisiche, che la scritta «°C» abbia un certo significato, eccetera.

Sembra, però, che il ragionamento a sistema riguardi la mia credenza sulla temperatura, non la verità, che la temperatura sia di 25 gradi è indipendente dal funzionamento del termometro; la coerenza è certo un ottimo criterio di giustificazione, ma pare inadatto come criterio di definizione della verità. Come aveva notato Bertrand Russell, una certa proposizione può essere coerente con certi sistemi e incoerente con certi altri, ma nel coerentismo ciò equivale a dire che è vera per certi sistemi e falsa per certi altri, quando è ovvio che una proposizione o è vera o è falsa, e non tutte e due le cose (Russell 1907).

A Russell si può rispondere che, proprio perché la verità è coerenza, non è vero che una proposizione o è vera o è falsa. Il coerentismo non accetta una assoluta non-contraddittorietà indipendente dal sistema di riferimento; non nega che una frase come «piove e non piove» deve essere falsa, ma accetta che «piove» e «non piove» possono essere entrambe vere per diverse persone o società con diversi sistemi di credenze. Il problema di questa risposta è che si può continuare a sostenere che ogni criterio diverso dal corrispondentismo sia *indicativo di verità*, ma non può essere la definizione di verità, proprio perché la relatività non è contemplata come una componente del comune concetto di verità. Dire che non è vero che una proposizione è o vera o falsa vuol dire *ridefinire* a piacere la verità, non scoprirne la definizione.

Com'è successo che la teoria coerentista sia passata da criterio di giustificazione a teoria della verità? È un evento legato allo sviluppo dell'idealismo, probabilmente risalente a Hegel ma che trova le sue radici già nel razionalista Spinoza (Walker 1989). Gli idealisti negano l'esistenza di fatti indipendenti dall'attività mentale, il loro ritratto del mondo è quello di un insieme di stati mentali; ne consegue che non ha senso cercare il fondamento della verità di una proposizione in qualcosa di esterno agli stati mentali, la verità deve essere la coerenza tra diversi stati mentali, nello specifico tra diverse credenze. Perciò, per questi filosofi, la coerenza non è solo indicativa di verità, ma verità punto, e l'obiezione di Russell non li raggiunge.

Anche la filosofia di Quine (2008) è molto vicina a questa posizione, è una sorta di coerentismo globale e radicale.

In modo direttamente opposto a idealisti come Fichte, Quine pare sospettare che non vi sia altro che conoscenza *a posteriori*. Le osservazioni empiriche che hanno posto le fondamenta della fisica quantistica sono fortemente controintuitive, spingono a conclusioni molto lontane dalla comune concezione del mondo quotidiano. Si poteva pensare che osservazioni così sfacciatamente contrarie al *senso comune* non possono essere corrette, e invece teorici come Schrödinger hanno interpretato la scoperta come necessità di rivedere la logica e rifiutare alcuni assiomi che fin ad allora erano stati considerati indubitabili. Altro esempio: la geometria euclidea si basa su assiomi splendidamente intuitivi, nessuno dubita che due rette parallele non si incrociano mai, non è neppure immaginabile diversamente, e invece gli sviluppi della fisica degli scorsi due secoli hanno portato a concludere che, qualsiasi sia la geometria alla quale risponde l'universo, deve essere non-euclidea. Pare davvero che tutto sia empirico, che non vi sia alcun fondamento indubitabile *a*

priori, man mano che la scienza produce nuove osservazioni tutte le conoscenze possono cambiare e adattarsi per formare un insieme coerente.

Questo è uno dei riflessi epistemologici della teoria della traduzione quineana: come non posso determinare il riferimento di *gavagai* se non in relazione al linguaggio che già parlo, così non posso determinare la verità di una qualsiasi affermazione se non in relazione alle altre credenze che già ho; come una parola non ha senso se non in una frase e una frase non ha senso se non in un linguaggio, così una affermazione scientifica sul mondo non è comprensibile se non in relazione ad un sistema di credenze già accettato.

Ciò che rende questa non solo una teoria della giustificazione, ma anche una teoria della verità, è che per Quine non c'è una traduzione corretta *di fatto*: i fatti esistono ma il riferimento delle parole non è determinato, è *inscrutabile*, non posso determinare la verità delle mie credenze solo in base all'esistenza dei fatti ai quali si riferiscono perché non c'è un riferimento oggettivo.

La tesi antirealista 3: il pragmatismo

La teoria della verità pragmatica sta nel chiarire il concetto di verità usando la massima pragmatica; la conclusione alla quale si arriva è spesso espressa con questo aforisma: è vero ciò che è utile.

Appena lo si sente sorgono le facilissime obiezioni di Russell (1994): l'utilità è indice di verità, ma non è né necessaria né sufficiente alla verità, è ovvio che possono esserci cose vere ma non utili, come utili ma non vere, la teoria delle stringhe potrebbe essere vera ma ciò non dipende in nessun modo dalla possibilità di trovare sue applicazioni di utilità pratica, il suo essere inutile non la renderebbe meno vera. Ovvio, ovvio: tutti i filosofi pragmatici accolgono subito questa obiezione, nessuno ha mai pensato di ridurre tutta la verità a quell'unico principio, nessuno ha mai separato la verità pragmatica dal realismo.

William James, uomo di esemplare tolleranza, ha sì sostenuto la teoria pragmatica della verità, ma sempre e solo in combinazione con coerentismo e corrispondentismo (e.g. James 2008). La mia impressione è che la sua sia una teoria *pluralista*, cioè credo riconosca diverse proposizioni come vere in diversi modi: ci sono affermazioni *fattualmente vere*, altre *coerentemente vere*, e infine alcune *pragmaticamente vere*.

James, usando la massima pragmatica, si chiede quale sia il vantaggio dell'aver credenze vere e in che modo questa verità debba manifestarsi nell'esperienza. Questo può portare a interpretare la teoria pragmatica come una *etica della credenza*: cioè si possono individuare idee che, anche non fossero fattuali, *sarebbe bene* credere vere comunque. Io potrei avere un tumore; le ricerche potrebbero dimostrare che credere in Dio aumenta la probabilità della remissione spontanea, che sia per miracolo o per qualche particolare processo neuroendocrinoimmunologico è irrilevante; fosse così, allora io farei bene a credere nella verità di Dio (cfr. James 1984).

Un approccio etico alla credenza come questo non è antiscientifico perché la credenza dipende sempre e comunque da fatti verificabili tramite l'esperienza sensoriale: nell'esempio che ho inventato il credere in Dio dipende dalla verifica empirica di un certo effetto statistico. Rimane però uno scarto tra questa verità pragmatica e il tipo di verità di cui normalmente si parla: dire che «sono giustificato a credere che l'esistenza di Dio sia un fatto» non implica che tale esistenza sia davvero fattuale.

Peirce (2011, e.g. 5.438 e ss.) infatti focalizza la sua teoria su altri aspetti: per lui la verità è la concordanza di una proposizione con il limite ideale verso il quale una infinita investigazione scientifica può tendere, è ciò su cui l'attività di ricerca concorderà o ciò che non riuscirà mai a rifiutare pur proseguendo infinitamente. Peirce non intende dire che questa *teoria definitiva* a cui può arrivare un'infinita investigazione scientifica sia perfetta, al contrario afferma che l'inaccuratezza, l'incompletezza di questa teoria è parte della sua verità, è vera anche *perché* afferma la sua propria imperfezione.

La teoria pragmatica avrebbe potuto essere coerente con il corrispondentismo; se non lo è, ciò è dovuto al diverso modo in cui le due prospettive interpretano lo scopo della teoria della verità. Per i corrispondentisti la teoria della verità è *descrittiva*, non *normativa*, deve dire cosa si intende con «verità» e non cosa *si dovrebbe* intendere, solo la teoria della giustificazione è normativa. Il pragmatismo rifiuta esplicitamente questa divisione: per Peirce la logica è normativa e il vero è un tipo di valore, il vero non è una componente della conoscenza, è ciò che la logica cerca di conoscere. Peirce guarda con sospetto alle definizioni realiste di verità, non gli piace il fatto che la descrivano come irraggiungibile, preferisce una teoria normativa in cui il vero è sempre *scopribile in linea di principio e inattaccabile dal dubbio*.

Ci sono prove contro l'antirealismo: il paradosso di Fitch

Secondo un celebre teorema intriso di pessimismo, le scoperte scientifiche non diventano mai note con il nome del loro vero scopritore (Stigler 1999). Questo vale anche per il paradosso di Fitch, pubblicato sì da Fitch (1963) ma su suggerimento di Church (Salerno 2009), uno dei più grandi matematici e logici del secolo scorso.

Se io nego che la verità stia nei fatti, ma sia bensì dipendente da me e dalla mia conoscenza, deduco che non ci può essere una verità inconoscibile: questo è il principio base dell'epistemologia antirealista ed è accettato da molte teorie epistemiche. Fitch analizza formalmente questo principio tramite la *logica epistemica*, una logica che permette di dare una rappresentazione in linguaggio formale, matematico, alle teorie della conoscenza: dimostra così che l'antirealismo implica assurdità che portano a dubitare della sua sensatezza.

Saper interpretare la scrittura logica è fondamentale per la filosofia analitica; io però voglio che ciò che dico sia comprensibile a tutti, e per spiegare la logica dovrei scrivere tutto un altro libro; e non ha neanche senso che lo faccia io visto che ne esistono già molti migliori di qualsiasi cosa possa produrre. Perciò, quando devo discutere un teorema originariamente scritto in logica simbolica, e non è troppo lungo o complesso, presenterò le formule (a volte in forma semplificata, saltando qualche passaggio) introducendole con una loro traduzione in lingua naturale. In ogni caso, anche nella discussione, cercherò di ridurre al minimo il formalismo.

Per ciò che segue, basti sapere che p indica un fatto qualsiasi espresso da qualsiasi proposizione (ad esempio il piovere), K indica il conoscere ciò che segue (Kp , so che piove), \rightarrow indica un'implicazione ($p \rightarrow b$, se piove allora mi bagno), \wedge indica la congiunzione «e» ($p \wedge b$, piove e mi bagno), \neg la negazione «non» ($\neg p$, non piove), \Diamond la possibilità ($\Diamond p$, può piovere). Per esempio, la formula $\neg \Diamond p \rightarrow \neg Kp$ si legge «se una cosa non è possibile, allora non la conosco».

Dunque, Fitch ragiona così:

Assiomi:

È chiaro, per la definizione di conoscenza, che se conosco un qualche fatto p questo deve esistere, p deve indicare una proposizione vera. Lo scrivo così:

$$\mathbf{A1} \quad Kp \rightarrow p$$

Se io conosco la congiunzione di due fatti, come «so che: piove e la strada è bagnata» saprò uno e anche l'altro, saprò che piove e saprò che la strada è bagnata:

$$\mathbf{A2} \quad K(p \wedge q) \rightarrow (Kp \wedge Kq)$$

Secondo l'antirealista, ogni verità è conoscibile. Ecco la sua tesi:

$$\mathbf{A3} \quad p \rightarrow \Diamond Kp$$

Infine, secondo Fitch l'antirealista, dalla dimostrata falsità di una proposizione, deduce l'impossibilità di quel fatto:

$$\mathbf{A4} \quad \neg p; \neg \Diamond p$$

Fitch dimostra che, se una proposizione è vera, allora è non solo conoscibile, ma *conosciuta*:

Teorema: $p \rightarrow Kp$

Dimostrazione:

Suppongo di sapere di non conoscere attualmente una qualche verità p . Questo lo scrivo così:

$$\mathbf{P1} \quad K(p \wedge \neg Kp)$$

Poiché conosco la congiunzione P1, per A2 conosco sia una cosa che l'altra, conosco sia che si dà il fatto p sia che non lo conosco, cioè:

$$\mathbf{P2} \quad Kp \wedge K\neg Kp$$

In virtù di A1, se è vero che conosco di non conoscere un fatto, allora è vero che non lo conosco, quindi posso riscrivere P2 così:

$$\mathbf{P3} \quad Kp \wedge \neg Kp$$

il che è contraddittorio, non posso conoscere e non conoscere una cosa allo stesso tempo; quindi la supposizione deve essere falsa e devo negarla, non è vero che io so che esiste un qualche fatto che non conosco:

$$\mathbf{P4} \quad \neg K(p \wedge \neg Kp)$$

Che tramite A4 implica l'impossibilità di conoscere ciò:

$$\mathbf{P5} \quad \neg \Diamond K(p \wedge \neg Kp)$$

Suppongo allora che, *a mia insaputa*, esista un qualche fatto che io non conosco, un'ipotesi molto modesta:

$$\mathbf{P6} \quad p \wedge \neg Kp$$

Sulla base di A3, come vuole l'antirealista, se P6 è una verità io posso conoscerla, è conoscibile che esiste un qualche fatto che non conosco, quindi scrivo:

$$\mathbf{P7} \quad \diamond K(p \wedge \neg Kp)$$

il che è però assurdo data P5, quindi la mia supposizione è falsa e devo negarla, non esiste un fatto che non conosco:

$$\mathbf{P8} \quad \neg(p \wedge \neg Kp)$$

e questo in logica classica è equivalente a dire che, se un fatto è vero, allora lo conosco:

$$\mathbf{C} \quad p \rightarrow Kp$$

QED

Quod erat demonstrandum, come volevasi dimostrare... ma è sbalorditivo! Se ogni verità è conoscibile, allora ogni verità è già conosciuta da qualcuno!

È così assurdo che deve esserci qualcosa di sbagliato. Ma cosa? La risposta più ovvia è la tesi antirealista stessa: non è possibile che ogni verità sia conoscibile, e quindi devono esserci verità indipendenti dalla mente, e quindi fatti. E quindi non è neanche un paradosso, ma un risultato razionale.

Argomentazioni di questo tipo sono chiamate *slingshot argument*, argomentazione-fionda: come il piccolo Davide ha sconfitto il gigante Golia semplicemente tirando un sasso, così un colosso filosofico come l'antirealismo è abbattuto da un ragionamento quasi banale (il nome viene da Barwise & Perry 1981). L'antirealismo però ha avuto tanti grandi sostenitori, tra i quali Michael Dummett (1986), uno dei più grandi filosofi inglesi e raffinato cultore della logica: possibile che non si siano mai resi conto di nulla? Ovviamente no.

Una delle tre regole del pensiero, una di quelle regole considerate l'innegabile base della razionalità, è quella del *terzo escluso*: o una affermazione è vera o lo è la sua negazione, o è vero che piove o è vero che non piove. Tale principio va a braccetto con quello della *doppia negazione*: due negazioni fanno un'affermazione, «non è vero che non è vero» equivale a dire «è vero». Sono questi principi che, nella dimostrazione di Fitch, permettono di dedurre, dal fatto che non esiste verità che non sia conosciuta, il fatto che ogni verità è conosciuta (non posso cioè passare da P8 a C).

Ma (c'è un ma) perché uno dovrebbe credere che o è vera un'affermazione o lo è la sua negazione? Ciò ha senso solo se la verità sta nei fatti: in tal caso è logico che il fatto affermato o si dà o non si dà, o la frase afferma un fatto che esiste o è un fatto che ciò che afferma non esiste. Ma se la verità non sta nei fatti, come vogliono gli antirealisti, allora non è necessariamente così: un antirealista può negare il principio del terzo escluso, e di conseguenza la doppia negazione; e se lo fa, dal fatto che non esiste verità che non sia conosciuta non si può più dedurre che ogni verità è conosciuta, lo si può negare senza contraddizione.

La logica *intuizionistica*, che nega il terzo escluso, permette di fare ciò. È una logica idonea a rappresentare il ragionamento quando si suppone la verità come costruita, non dipendente dai fatti. Se la si accetta il paradosso di Fitch non emerge (Williamson 1982).

D'altro canto, questo non vuol dire che in logica intuizionistica non sorgano altri, nuovi paradossi; inoltre, cambiando logica si cambiano le basi stesse del ragionamento, il che è possibile ma davvero radicale. Gli antirealisti che non vedono nell'intuizionismo una soluzione preferiscono sostenere semplicemente che il paradosso fraintende il principio di conoscibilità, che Fitch non rappresenta ciò che davvero gli antirealisti credono. Dummett, per esempio, sostiene che un antirealismo *globale* sia impossibile, l'antirealismo ha la pretesa di valere solo per alcuni settori, quindi non è vero che nell'antirealismo *ogni* verità è conoscibile (Dummett 2006). Ne consegue, per esempio, che dal principio di conoscibilità non si può implicare che posso sapere che si dà un fatto che non conosco. Il problema di questa soluzione è che la teoria creata si chiama antirealismo, ma non è antirealista, è un pluralismo come quello di James, in cui alcune verità sono fattuali e altre no.

Non ci sono prove a favore dell'antirealismo 1: terre gemelle e cervelli in vasca

Putnam è un filosofo per me fondamentale, credo sia stato il più grande al mondo dopo la dipartita di Quine, col quale aveva spesso collaborato e al quale è andato spesso contro. Come Dante ha scelto Virgilio per guidarlo attraverso l'inferno, io sceglierei Putnam. Il primo motivo è che sono affascinato da quei pensatori che uniscono matematica e linguaggio, queste dimensioni così distanti dell'intelligenza: e Putnam, come Quine, è stato anche un matematico di rilievo, autore di teoremi importanti e algoritmi ben noti a chi si occupa di informatica (si veda Davis & Putnam 1960). Il secondo motivo è che era un uomo totalmente libero da bias e pregiudizi, per nulla affezionato alle sue idee, sempre pronto a sottoporre le sue opere a profondissima autocritica. Per questo è un filosofo col quale non si può non andare d'accordo: in qualsiasi argomento abbia voluto dire la sua, prima ha sostenuto una certa tesi, e poi il suo opposto, cambiava di continuo opinione, così chi non si trova d'accordo con il Putnam di un certo periodo si troverà d'accordo con quello del periodo successivo o precedente.

Putnam, come Quine, è noto per l'essere un realista e un naturalista, eppure è proprio lui l'autore delle più potenti argomentazioni contro il realismo: la sua fiducia nella verità della scienza non si è sempre accompagnata all'idea che questa verità sia di tipo fattuale. L'argomentazione che voglio esporre ora è composta da due esperimenti mentali, il primo riguarda la natura del significato linguistico e giustifica una particolare teoria semantica (Putnam 1987), il secondo parte da questa teoria per trarre conclusioni antirealiste (Putnam 1985).

Alcuni astronomi hanno immaginato che, posto un universo infinito, andando avanti e avanti tra le galassie ad un certo punto si trovi un dopplegänger, una zona di universo in cui la materia è disposta esattamente così com'è disposta qui. Dovrebbe esserci, da qualche parte nell'universo, una Terra, la chiamerò Terra*, esattamente uguale a questa Terra, dove si trova un uomo* esattamente uguale all'uomo che sono io, disteso davanti ad un fiume* esattamente uguale al fiume davanti al quale sono disteso io... con un'unica differenza: la mia acqua ha formula chimica H₂O, mentre la sua acqua*, pur comportandosi in modo identico a quella che conosco, è caratterizzata da una diversa formula XYZ. Ora io punto il dito verso il fiume e dico «acqua!», nello stesso istante l'uomo* fa lo stesso e dice «acqua!» proprio come me. Ma: qual è il contenuto delle parole? Quali enti del mondo rappresentano le cose che stiamo dicendo?

Non ho dubbi che la mia parola «acqua» si riferisca all'acqua di formula H_2O , mentre nel caso dell'uomo* il riferimento sia l'acqua* di formula XYZ. Ma questo è sorprendente! Io e l'uomo* siamo in tutto e per tutto uguali, gli stati mentali sono identici; può darsi che nessuno di noi due conosca la formula chimica della sostanza che ha davanti, o può anche essere che l'uomo* creda erroneamente (come credo io, giustamente) che la formula del liquido che ha davanti sia H_2O invece di XYZ, ma questo non cambia, l'uomo* non può riferirsi all'acqua del mio pianeta, sia perché non l'ha mai vista, sia perché indica la sua e non la mia sostanza.

Parto con l'astronave diretto verso la Terra*, atterro, faccio l'analisi chimica dell'acqua* e concludo «su questo pianeta, gli abitanti chiamano *acqua* il liquido XYZ, e non l' H_2O ». Che succede? Ho riconosciuto che quella parola, pur uguale alla mia, ha un'estensione differente, e quindi un diverso significato, insomma che è un'altra parola, indipendentemente dagli stati mentali dell'uomo*. Questo è simile al caso in cui io, andato dal mio meccanico di fiducia, gli dica «forse la macchina ha un problema con lo spinterogeno»: io non ho idea di cosa sia lo spinterogeno, cosa indichi questa parola non dipende affatto da ciò che io penso in merito, accetto invece che si riferisca a ciò che il meccanico ritiene indichi.

Ciò che questo esperimento mentale vuole sostenere è che non esiste alcuno stato mentale interno che possa determinare il significato delle parole, ma può essere determinato da elementi esterni. Penso ad un bambino che, dopo la prima lezione di scienze, dica «quando dicevo acqua intendevo il liquido H_2O , anche se non lo sapevo»: il bambino non aveva neppure mai pensato a cosa indicasse quella parola, accetta che il significato non sia determinato da fatti a sé interni come i suoi stati mentali, accetta invece che il significato dipenda dal contesto in cui agisce, da elementi esterni.

Questa appena esposta è la teoria dell'*esternalismo semantico*. Putnam la collega alla *teoria causale del riferimento*: cioè l'uomo* quando dice «acqua» si riferisce all'acqua*, e non all' H_2O , perché quella è la sostanza che ha *causato* l'uso di quella parola, l'uomo* non può riferirsi all'acqua del mio pianeta con la quale non ha alcun rapporto causale, non c'è alcuna catena causale che connetta l'uso della sua parola alla mia acqua. Queste teorie sono presupposti del secondo esperimento mentale.

Per quel che ne so, io potrei essere un *cervello in vasca*. Cioè, potrei non avere il corpo che vedo, potrei non abitare il mondo che vedo, bensì, come nel film *Matrix*, potrei essere collegato ad un computer che invia dei segnali al mio cervello e lo fa reagire come se visse tutto ciò, anche se in realtà il mondo è senza case, senza alberi, senza corpi in carne e ossa, ed è dominato dalle macchine. Questo contesto distruggerebbe la mia capacità di conoscere il mondo ma è coerente con il realismo in quanto prevede una scissione tra il mondo come lo esperisco e il mondo così com'è.

Ora dico «ho una incredibile intuizione: io sono un cervello in vasca!» Mi aspetto che questo volo verso i più alti reami della coscienza mi doni incredibili superpoteri, come Deadpool che sapendo di esistere solo all'interno di un fumetto può fare quello che vuole; e invece l'unico effetto è quello di farmi internare in qualche ospedale psichiatrico. Che senso ha il mio dire «io sono un cervello in una vasca»? A quali oggetti si riferiscono le parole «cervello» e «vasca»?

Se io sono un cervello in vasca, la mia esperienza degli oggetti non è una rappresentazione di oggetti reali, è solo l'effetto di un certo codice del computer che gestisce gli impulsi che viaggiano verso i miei nervi. Forse nel mondo esterno c'è qualcosa che assomiglia ad un cervello, ma io non l'ho mai visto, ho visto solo i cervelli che si danno nel mondo della mia esperienza. Per la teoria causale del riferimento, quando io dico «cervello» non posso riferirmi a qualche oggetto del

mondo reale col quale non ho mai avuto contatti, al massimo posso riferirmi agli impulsi che arrivano ai miei nervi, al codice del computer che genera il mondo che vedo. È dunque certo che io non sono un cervello in vasca perché io non sono un'esperienza, un codice, nulla di ciò a cui le mie parole possono riferirsi; posso immaginare di essere un cervello in vasca, ma immagino qualcosa di falso.

Lo ripeto nello stile schematico della Terra gemella. Esiste un mondo*, esterno al mondo della mia esperienza. In questo mondo*, io sono un cervello* in una vasca* collegato ad un computer che genera il mondo (senza asterisco) della mia esperienza; in questo mondo ci sono cervelli e vasche (pure senza asterisco). Io posso affermare «sono un cervello in vasca!», ma non potrò mai affermare «sono un cervello* in vasca*!»: come l'uomo* della Terra* si riferisce all'acqua* indipendentemente da ciò che pensa d'essa, così quando un cervello in vasca dice «vasca» non può riferirsi alla vasca* del mondo reale indipendentemente da ciò che può pensare; proprio per questo posso dire di non essere un cervello in una vasca, perché non posso riferirmi a cose del mondo*.

Se io sono un cervello in vasca, sono come un grammofono: il grammofono emette suoni di parole sensate, ma non può riferirsi a nulla di reale. Sono come un computer, che sembra parlarmi anche se le sue parole non hanno alcun riferimento; sono come una formica che traccia una scritta sulla sabbia, anche se non ha alcuna idea di quello che sta facendo. Se gli alberi sparissero di colpo dal mondo, il grammofono continuerebbe a gracchiare discorsi sugli alberi, i computer continuerebbero a scrivere le frasi sugli alberi che i loro algoritmi creano, la formica continuerebbe a scrivere «albero» sulla sabbia; e lo stesso farei io, se sono un cervello in vasca, che parlo di alberi anche se non esiste proprio alcun albero.

Perché tale esperimento è problematico per il realismo?

- P1** Se il realismo è vero, allora è possibile che io sia un cervello in vasca.
- P2** Se sono un cervello in vasca, non ho la facoltà di esprimere significato, non posso riferirmi ai fatti del mondo, e quindi il corrispondentismo è falso.
- P3** Ma se è così, allora io non posso essere un cervello in vasca.
- C** Quindi è impossibile che io sia un cervello in vasca, quindi il realismo è falso.

Questo non vuol dire che il mondo esterno non esista, vuol dire che il mondo non può avere una struttura indipendente dal pensiero, come in Kant il mondo è inconcepibile se non tramite gli schemi mentali.

Il primo punto debole dell'argomentazione sta nell'idea che il caso dei cervelli in vasca implichi la falsità del corrispondentismo: se l'esternalismo semantico e la teoria causale del riferimento vengono rifiutati anche questo punto crolla. C'è qualcosa di non convincente nell'esperimento della Terra gemella?

Ci sono io, uomo della Terra, e c'è l'uomo* della Terra*, entrambi con un bicchiere in mano che diciamo «questa è acqua», ma in un bicchiere c'è H₂O e nell'altro XYZ. Putnam, ispirato da Kripke (1999), tratta «acqua» come se fosse un nome proprio della sostanza H₂O. Io posso dire «Einstein sarebbe potuto essere un rapper»: Einstein è un nome proprio e, per *stipulazione*, indica sempre lo stesso individuo, non importa quali siano le caratteristiche che vengono alterate, non posso dire «quel rapper non sarebbe lui». Similmente, «acqua» è come fosse un nome proprio che, per stipulazione, indica sempre H₂O, sia sulla Terra che sulla Terra*: in tale teoria posso dire «questa

acqua sarebbe potuta essere potabile», ma dire «questa acqua sarebbe potuta non essere H₂O» sarebbe come dire «Einstein sarebbe potuto non essere Einstein», inaccettabile.

I nomi che mostrano questo comportamento sono detti *designatori rigidi*. Non c'è dubbio che i nomi propri siano di questo tipo, ma perché un nome di sostanza com'è «acqua» dovrebbe funzionare nello stesso modo? I *designatori flaccidi* indicano enti diversi in diversi contesti, mentre il rigido «Trump» indica sempre lo stesso individuo, il flaccido «il presidente» indica una persona diversa a seconda dello Stato e dell'anno di riferimento: perché «acqua» non potrebbe funzionare così? Non è forse vero che se H₂O e XYZ fossero due liquidi dello stesso pianeta, chiamati entrambi «acqua» dalla popolazione, sarebbero entrambi nell'estensione della parola?

Se io e l'uomo* della Terra* abbiamo un concetto *vago* di acqua, come generica sostanza acquosa, allora potremmo benissimo considerare sia H₂O che XYZ come acqua, solo dopo aver scoperto la composizione chimica si potrà decidere se indicare con quel nome solo una delle due o continuare a indicare entrambe. In questo caso, è vero, non si può scoprire che l'acqua è H₂O, quando si scopre la composizione chimica si crea anche una nuova definizione di «acqua»: ma se per Putnam questo è indesiderabile, per Searle si tratta di una pratica normale (Searle 1983).

Il secondo punto debole dell'argomentazione sta nel derivare l'impossibilità dei cervelli in vasca dalla falsità del corrispondentismo. Anche accettando che, se si è dei cervelli in vasca, allora il corrispondentismo è falso, da ciò non segue che i cervelli in vasca sono una impossibilità metafisica.

Putnam dice che, se io sono un cervello in vasca, allora «cervello» e «vasca» non possono riferirsi a cervelli e vasche, quindi «sono un cervello in una vasca» è falso. Ciò che avrebbe dovuto dire però è che, se io sono un cervello* in vasca*, allora «cervello» e «vasca» non possono riferirsi a cervelli* e vasche*: in questo modo si vede che io posso davvero escludere di essere un cervello in vasca, ma non posso escludere anche di essere un cervello* in vasca*; l'esserlo può risultarmi inconcepibile e inesprimibile, ma ciò non vuol dire che sia impossibile (Khlenzovs 2021). Sì, se sono un cervello in vasca le mie parole non si riferiscono ad alcun oggetto esterno, e se lo fanno come io credo allora non sono un cervello in vasca; di fatto però non posso escludere che le mie parole non si riferiscano ad alcunché di esterno.

Questo rimane problematico per il realismo: se non posso escludere che le mie parole non si riferiscono a nulla di esterno, allora non posso essere giustificato a credere che lo facciano, anzi dovrei credere che si riferiscano direttamente alle mie esperienze. Questo si lega ad un problema aperto complesso: com'è che le parole si legano agli oggetti reali, com'è che possono riferirsi anche a cose inimmaginabili. Tornerò su questi punti più avanti.

Non ci sono prove a favore dell'antirealismo 2: la teoria dei modelli

L'argomentazione che esporrò ora costituisce uno dei passaggi più complessi e importanti di tutto il libro, e lo richiamerò spesso anche nei ragionamenti successivi.

Che cosa sto per scrivere? In sunto: se il realismo è vero, se il mondo è indipendente dagli schemi mentali e dalla conoscibilità, allora perfino la più perfetta teoria scientifica producibile dall'umanità potrebbe essere falsa (magari non lo è ma potrebbe); secondo Putnam, a causa dell'indeterminatezza della traduzione di Quine, *deve* esistere una interpretazione vera di tale teoria,

non è vero che può essere falsa, quindi non è vero che il mondo è indipendente dagli schemi mentali.

Scriverò questa argomentazione in due modi, come fa Putnam stesso: la prima volta da una prospettiva generale e tecnica facendo ricorso ad alcuni concetti logici e matematici, la seconda tramite un esempio specifico e concreto (Putnam 1977, 1980 e 1985).

Putnam, come Quine, si appoggia al settore matematico della teoria dei modelli. Quindi, in questo contesto: cos'è una teoria e cos'è un modello?

Io ho tre amici, Alessandra (a), Bruno (b) e Carlo (c). Alessandra e Bruno hanno completato un dottorato in storia del canto rinascimentale, mentre Carlo ha fatto una inutile laurea in ingegneria: Alessandra e Bruno sono storici (S), mentre Carlo non lo è (-S); schematicamente: «Sa, Sb, -Sc». Queste tre espressioni costituiscono la mia teoria sui miei amici: una *teoria* è semplicemente un insieme di espressioni.

Queste espressioni fanno affermazioni su certe entità del mondo che costituiscono l'estensione dei termini usati, cioè sono anche proposizioni: a indica Alessandra, b Bruno, c Carlo. L'*essere-storico* S pure si definisce tramite la sua estensione, cioè $S=[\text{Alessandra}, \text{Bruno}]$. Questa è una interpretazione possibile della teoria: una *interpretazione* è una funzione che assegna ad ogni termine una estensione (da una parte ci sono i miei simboli, dall'altra ci sono delle entità, la funzione traccia le linee tra le due).

Ora posso calcolare il valore di verità delle proposizioni: per esempio, «Sa» è vero se l'estensione di a è parte anche dell'estensione di S, cioè se Alessandra è una storica. Poste le estensioni viste, le tre proposizioni della teoria «Sa, Sb, -Sc» risultano tutte e tre vere. Una interpretazione che *soddisfa* tutte le proposizioni della teoria, cioè una funzione che assegna ad ogni termine una estensione tale da rendere tutte le proposizioni espresse vere, è un *modello*.

Ora spiego questa teoria a mia nonna che, non conoscendo i miei amici, fa confusione: pensa che a sia Bruno, b sia Carlo, c sia Alessandra, e che S voglia dire *essere-maschi*, e quindi l'estensione sia $S=[\text{Bruno}, \text{Carlo}]$. Stando così le cose, le proposizioni espresse da «Sa, Sb, -Sc» (che sono diverse da quelle che volevo esprimere io) risultano comunque tutte e tre vere; quindi esistono almeno due modelli diversi, almeno due interpretazioni tali che ogni proposizione espressa dalle frasi della teoria risulta vera. Il mio era il modello *inteso*, quello capito da mia nonna è un modello *alternativo*. Il modello inteso può essere distinto dagli alternativi ponendo dei *vincoli*: per esempio posso dire a mia nonna che S deve poter includere uomini, donne, ragazzi, ragazze, ma non bambini o bambine (visto che i bambini non hanno un dottorato: il più giovane dottore di sempre, Karl Witte, lo è diventato a 13 anni).

Adesso devo introdurre a grandissime linee alcuni teoremi utili alla riflessione sui modelli.

Kurt Gödel, il più grande logico di sempre, è noto soprattutto per i suoi teoremi di incompletezza, dei quali parlerò in metafisica; ora mi interessa invece il teorema di *completezza* per la logica del primo ordine (che è quella di Frege e Peirce). Da tale teorema si può dedurre che qualsiasi insieme di espressioni in logica del primo ordine che non sia contraddittorio ha un modello (Gödel 1999-2009, vol. 1). E tanto mi basta, per ora.

Un altro gigante della matematica è Georg Cantor. Cantor è famoso per il suo magnifico lavoro nella teoria degli insiemi e l'analisi del concetto di infinito in matematica: un tema affascinante,

emozionante, che rende la sua opera d'interesse anche per coloro che non si occupano di matematica ma la vogliono usare per discutere di filosofia o teologia (Cantor 2020).

Cantor ha avuto una intuizione profonda: l'infinito *assoluto* non è l'unico, ci sono più infiniti, e non tutti gli infiniti sono, per così dire, ugualmente infiniti. Voglio fare un esempio intuitivo anche se so che è sbagliato: tra 1 e 2 c'è un'infinità di altri numeri (1,1; 1,11; 1,111...), tra 2 e 3 pure c'è un'infinità di numeri, ma tra 1 e 3 c'è un'infinità di numeri più grande delle altre due. Ricordo che la prima volta che ho pensato a questa cosa non riuscivo a capacitarmene... E avevo ragione perché è sbagliata, *i numeri tra 1 e 2* e *i numeri tra 1 e 3* sono insieme con la stessa *cardinalità*. La cardinalità è il modo in cui si indica la dimensione di un insieme e quindi, per usare un'espressione molto impropria, anche diverse dimensioni dell'infinito: l'insieme *la mia famiglia* (io, i miei genitori, e il cane) ha cardinalità *finita* pari a quattro, *i numeri naturali* (1, 2, 3...) ha una cardinalità *infinita* esprimibile con un certo numero *transfinito* (nello specifico un numero detto *aleph-zero*, scritto \aleph_0) ma inferiore a quella dell'insieme infinito *i numeri reali* (1, -3, $\frac{5}{4}$...) espressa da un diverso numero *transfinito* (chiamato *il continuo*, \aleph_1 , o \mathfrak{c}); i numeri naturali sono infiniti, i numeri reali pure ma in modo più elevato.

Il fatto che un certo insieme, come *i numeri tra 1 e 2*, sia infinito non vuol dire che non sia *contabile*. Cantor definisce «contabili» quegli insiemi che possono essere messi in una relazione *uno-a-uno* con l'insieme dei numeri naturali, come se potessi tracciare linee per collegare ogni elemento dell'insieme contabile ad un numero naturale: per esempio, i numeri interi sono infiniti, ma contabili.

Altri due matematici, Skolem e Löwenheim, danno il loro nome ad un importante teorema. Come nel caso di Gödel, non voglio approfondirlo, mi concentro solo sugli aspetti rilevanti per l'argomentazione di Putnam: mi basta sapere che, in primo luogo, se un insieme di espressioni in logica del primo ordine ha un modello infinito, allora ha anche un modello che è contabile; in secondo luogo, se un insieme di espressioni in logica del primo ordine è contabile e ha un modello infinito di una qualsiasi cardinalità, allora ha un modello di ogni cardinalità infinita (Skolem 1977).

Tutto è pronto per la prima versione dell'argomentazione di Putnam.

Immagino di avere una bellissima teoria dell'universo, la teoria T, scevra da qualsiasi contraddizione ed esprimibile in logica del primo ordine. Ipotizzo anche che l'universo descritto da questa teoria sia infinito come la teoria stessa afferma, e costituisca quindi un insieme che contiene infiniti enti reali.

Posto ciò, tramite il teorema di Gödel so che T avrà un modello, una interpretazione tale che le proposizioni espresse dalla teoria siano vere. Tramite il teorema di Löwenheim-Skolem so che T avrà un modello di ogni cardinalità infinita, e quindi avrà anche un modello M di cardinalità uguale a quella dell'universo. Non so ancora, però, se gli elementi di M (le estensioni dei termini della teoria T) siano anche enti dell'universo: potrebbero benissimo essere solo numeri. Una conseguenza del teorema di Löwenheim-Skolem è che, se l'insieme degli elementi di M e l'insieme dell'universo hanno uguale cardinalità, allora esiste un modello anche ad estensioni invertite: cioè posso creare un nuovo modello W i cui elementi sono tutti gli oggetti dell'universo, di modo che in tale modello ogni espressione della teoria T è associata a sottoinsiemi di oggetti reali. È così possibile dire che la teoria T deve essere *vera nell'interpretazione W*.

In altre parole, se T ha le caratteristiche dette, allora esiste una funzione, o interpretazione (che è W), che assegna alle espressioni di T una estensione di enti dell'universo tale che ogni proposizione di T risulta vera relativamente alla tal funzione. In altre parole ancora: posta l'indeterminatezza del linguaggio vista con Quine, posto cioè che uno stesso insieme di frasi può esprimere più proposizioni, una teoria come quella descritta può essere sicuramente interpretata in modo tale da risultare vera.

Sorge un'obiezione istintiva: W non è il modello inteso, può essere una interpretazione alternativa, il fatto che la teoria sia *vera nell'interpretazione W* non vuol dire che sia vera *tout court*. Mi chiedo: come faccio a individuare il modello inteso? Pongo dei vincoli specificando delle condizioni di verità aggiuntive: per esempio, posso dire che «in questa stanza ci sono 25 gradi» è vero se il mercurio del termometro si espande fino al segno «25 °C», questo pone un vincolo che esclude tutte le interpretazioni in cui «ci sono 25 gradi» è vero anche se il mercurio non si espande fino al tal segno.

Epperò, una teoria come quella descritta è dotata di ogni virtù teoretica e può superare ogni test di questo tipo. Perché? Proprio perché risulta vera. La teoria può affermare al suo interno di essere semplice, coerente, utile, di rispettare il senso comune, e tutti questi possibili vincoli risulterebbero veri in quanto nell'interpretazione W ogni sua proposizione è vera: la teoria rispetta tutti i vincoli che sono esprimibili nel suo linguaggio. Quali altri vincoli potrei porre? Putnam risponde che non c'è alcun altro vincolo che sia razionale porre: anche se ponessi infiniti vincoli non arriverei mai ad escludere tutti gli infiniti modelli alternativi possibili, quindi anche dubitassi che W sia il modello inteso comunque non potrei distinguere quello inteso dagli altri, e quindi non c'è motivo per il quale W non debba essere considerato il modello inteso.

Passo ora alla seconda versione dell'argomentazione. Tale versione mostra in modo più esplicito come, posta una interpretazione della teoria in cui ogni espressione ha una certa estensione, è possibile rimescolare le estensioni senza cambiare i valori di verità delle proposizioni; e, in questo, evidenzia ulteriori collegamenti tra l'argomentazione di Putnam e la teoria di Quine.

Immagino vi siano tre tipi di universi possibili:

1. nel primo esistono gatti che stanno su tappeti, e ciliegie che stanno sugli alberi;
2. nel secondo esistono gatti che stanno sui tappeti, ma non ciliegie che stanno sugli alberi;
3. il terzo può essere qualsiasi altro universo di tipo diverso dagli altri due.

La mia teoria è «il gatto è sul tappeto». L'interpretazione standard è quella che tutti coloro che parlano italiano capiscono. Voglio però immaginare un diverso modello, immagino che un alieno che sta imparando la mia lingua la interpreti come «il gatto* è sul tappeto*». I nuovi concetti *gatto** e *tappeto** sono speciali, hanno un'estensione diversa da quella standard:

- «gatto*» significa proprio *gatto* nel caso del secondo universo, e significa *ciliegia* in tutti gli altri casi
- «tappeto*» significa proprio *tappeto* nel secondo caso, *albero* nel primo, e *protone* nel terzo

Per strano che possa sembrare, «il gatto* è sul tappeto*» è logicamente equivalente a «il gatto è sul tappeto»: se si dà uno si dà anche l'altro. Infatti:

1. nel primo universo «il gatto è sul tappeto» è vero, e «il gatto* è sul tappeto*» cioè «la ciliegia è sull'albero» pure lo è;

2. nel secondo universo «il gatto è sul tappeto» è vero, e «il gatto* è sul tappeto*» avendo lo stesso significato dell'altra pure lo è;
3. nel terzo universo «il gatto è sul tappeto» deve essere falso, e «il gatto* è sul tappeto*» cioè «la ciliegia è sul protone» pure deve esserlo.

Incredibile: il gatto è sul tappeto se e solo se il gatto* è sul tappeto*!

Sembra possibile porre un facile vincolo per distinguere il modello inteso: cosa sia l'*essere gatto* non varia, mentre l'*essere gatto** varia col contesto, varia con l'universo in cui mi trovo; posso quindi porre l'*invariabilità* come vincolo della teoria. Non è però un vincolo valido poiché anche «gatto» può avere una definizione contestuale. Per esempio, immagino che «la ciliegia è sull'albero» abbia come interpretazione alternativa «la ciliegia* è sull'albero*» dove:

- «ciliegia*» significa proprio *ciliegia* nel secondo universo, e significa *gatto* in tutti gli altri casi;
- «albero*» significa proprio *albero* nel secondo universo, *tappeto* nel primo, e *protone* nel terzo

Ora «ci sono gatti su tappeti e ciliegie sugli alberi» e «ci sono gatti* su tappeti* e ciliegie* sugli alberi*» sono uguali, posso combinare parole con e senza asterischi a piacere, magari per dare una definizione contestuale di «gatto»:

- «gatto» significa *gatto** nel secondo universo e *ciliegia** in tutti gli altri

Un altro vincolo può riguardare il valore di verità delle parole in contesti verbali: se io sto osservando un gatto, la frase «io vedo un gatto» è sicuramente vera, mentre non posso giudicare vera o falsa «io vedo un gatto*» finché non determino in quale universo mi trovo in quanto nel primo è falsa mentre nel secondo è vera. Ma neanche questo vincolo è risolutivo: basta dare una reinterpretazione del verbo, *vedo**, e poi dare a «vedo» e «vedo*» definizioni contestuali tali che i valori di verità delle due frasi siano gli stessi in tutti gli universi. Si può andare avanti così all'infinito con ogni frase, quindi è una strategia che non funziona.

Passo allora agli stati mentali. Quando io dico «il gatto è sul tappeto» penso proprio lo stesso, ma quando dico «il gatto* è sul tappeto*» penso «la ciliegia è sull'albero»; quando guardo un gatto penso all'*essere gatto* e non all'*essere gatto**, alla *gattità* e non alla *gattità**. Posso dunque usare i miei stati mentali come vincolo per distinguere tra le due interpretazioni? Il Kripkenstein si fa questa domanda in merito al *rule-following paradox* e si risponde di no: il pensiero, le immagini mentali, le sensazioni, qualsiasi circuito neurale possa rappresentare il pensiero del *gatto*, tutto ciò deve essere a sua volta interpretato, come posso reinterpretare «vedo» in «vedo*» così posso reinterpretare i miei stati mentali o neurali annullando questa differenza tra i due casi. Se si accetta l'esternalismo è ancora più ovvio: il fatto che l'uomo* della Terra* dica «acqua!» pensando all' H_2O in nessun modo cambia il fatto che la sua parola indica invece la sostanza XYZ, similmente ciò che io penso dicendo «il gatto è il suo tappeto» in nessun modo può fissare la sua interpretazione.

Vedo che «il gatto è sul tappeto» ha più interpretazioni, diverse in quanto a significato ma equivalenti in quanto a valore di verità; ho fatto l'avvocato del diavolo cercando di dimostrare che si possono escludere le interpretazioni alternative, che è possibile distinguerle da quella intesa, ma non ci sono riuscito. Se in un certo istante il numero dei gatti e il numero delle ciliegie fosse uguale potrei addirittura scambiare gatti e ciliegie e i valori di verità della teoria non cambierebbero. E questo è l'ultimo passaggio visto nella prima esposizione quando si è usato il teorema di Löwenheim-Skolem per scambiare gli elementi di M con gli enti dell'universo in virtù della loro uguale cardinalità.

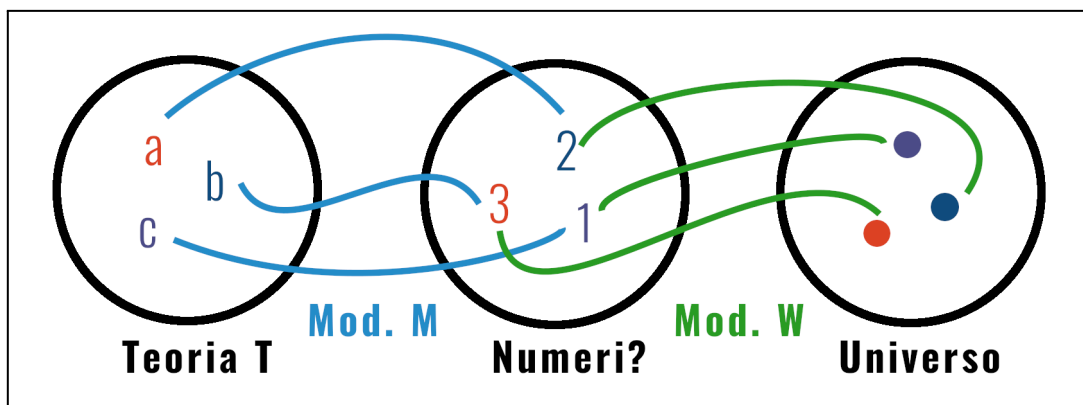


Figura 1

Questa è una indeterminatezza profonda che riguarda addirittura le mie percezioni: ogni esperienza ha infinite cause possibili, infiniti diversi ritratti del mondo possono tradursi nelle stesse esperienze. Non importa quanto eccellenti siano le teorie scientifiche e quanto restrittivi i vincoli, posso anche specificare il valore di verità che deve avere ogni singola proposizione in ogni possibile circostanza, comunque vi saranno interpretazioni alternative che danno un diverso ritratto del mondo. Da qui discende la *relatività ontologica* di Quine.

Si può essere tentati (dal demonio) a considerare questa *relatività ontologica* come un *relativismo* antiscientifico: poiché non c'è un fatto in merito al riferimento, tanto vale fare il passo successivo e affermare che non c'è alcun fatto in generale, ci sono solo interpretazioni, anche ciò che le teorie scientifiche dicono deve essere del tutto soggettivo.

Né Quine né Putnam hanno mai sostenuto qualcosa di simile. Al contrario, Putnam mette in chiaro che se la specie umana non fosse pervenuta a intuizioni e credenze in gran parte vere allora la sua sopravvivenza sul pianeta Terra sarebbe inspiegabile; corollario, se le credenze scientifiche sviluppate nei secoli non fossero in gran parte vere allora il successo della scienza nel prevedere l'evoluzione del mondo sarebbe inspiegabile (Putnam 1985). Se sono qui a scrivere è perché esiste un mondo esterno conoscibile e conosciuto. La relatività ontologica non sta nell'inesistenza di un mondo oggettivo, ma nel fatto che ci sono più modi di conoscere e interpretare questo mondo.

Prendo due amici vicentini. Notoriamente, i vicentini sono golosi di carne di gatto; loro però sono vegani, e preferiscono le ciliegie. Il primo amico ha un ritratto del mondo che prevede *gatti**, *tappeti**, *alberi*, *ciliegie* (la chiamo *interpretazione 1*), mentre il ritratto dell'altro prevede *gatti*, *tappeti*, *alberi** e *ciliegie** (*interpretazione 2*). Il primo amico dice al secondo «andiamo a raccogliere delle ciliegie!», e il secondo capisce «andiamo a raccogliere* delle ciliegie*!», cioè «andiamo ad accarezzare dei gatti!»; fortunatamente, *ciliegia* e *ciliegia** sono concetti collegati alle stesse esperienze soggettive, cioè il secondo amico, lì dove il primo vede delle ciliegie, crede di vedere ciliegie* cioè gatti, e quando il primo ha la sensazione di raccogliere ciliegie, al secondo pare di avere la sensazione di accarezzare un gatto. Anche se il loro modo di intendere il mondo è così diverso, il tipo di comportamento che ne consegue è identico: andranno entrambi a raccogliere ciliegie.

Può sembrare la descrizione di un nuovo tipo di follia, ma è solo una reinterpretazione dell'esempio di Quine: un cuoco mi dice «vammì a prendere un *gavagai*», e lui intende *coniglio*, io invece intendo *coniglio**, cioè sezione tridimensionale dell'ente quadridimensionale noto come coniglio, ma sia *coniglio* che *coniglio** sono associati all'esperienza soggettiva dello stesso ente (dove

gli altri vedono un coniglio io vedo una sezione tridimensionale), e così, indipendentemente da come io lo concettualizzi, porterò al cuoco proprio ciò che vuole.

Se io ho una credenza palesemente errata come «è bene mettere la testa tra le fauci di un leone mentre sbadiglia», che questo sia un errore sarà presto visibile a tutti; ma se io ho due interpretazioni alternative con identici valori di verità e *comportamentalmente indistinguibili*, come nel caso dei due amici, allora non ho motivo di dire che una delle due sia sbagliata. La *verità-in-1* e la *verità-in-2* sono ugualmente valide: non c'è nulla, nessun comportamento o vantaggio evolutivo, che possa portarmi a giustificare un'unica interpretazione, un unico modo di legare le espressioni agli oggetti esterni.

La verità, in tale argomentazione, deve dunque dipendere da un certo punto di vista interno al mondo di cui si parla (e perciò non può stare nei fatti). Il motivo per cui Putnam ritiene i cervelli in vasca una impossibilità metafisica, pur essendo concepibili come una possibilità fisica, è che in un mondo di soli cervelli in vasca non esiste alcun punto di vista se non quello dei cervelli, e dunque non esiste alcuna verità al di fuori di essi. Opinioni simili si trovano in vari grandi filosofi del secolo scorso quali Sellars (2004), Rorty (2004) e Dummett (2008); Dummett, in particolare, arrivò a dire che, se è possibile una descrizione scientifica totalmente oggettiva del mondo, allora deve esserci un Dio, perché l'unico punto di vista in grado di rendere vere certe proposizioni è il Suo.

Mi piace interpretare questa argomentazione come riflesso di un *criticismo neopragmatico*, una visione kantiana, antirealista ma moderata: gli enti esterni esistono, ed è possibile conoscerli e parlarne, ma non in modo indipendente dalle categorie mentali. Putnam si chiede come sia possibile che le categorie mentali finiscano per coincidere con le categorie del mondo: forse che i gatti nascono già in una certa categoria, distinta dalla categoria delle ciliegie, a prescindere dal mio categorizzarli come gatti? Si risponde che in realtà il mondo non ha alcuna struttura o divisione per categorie, il mondo è un ammasso indistinto di *roba*, è l'uomo che lo spezzetta in oggetti imponendo la sua struttura. Non c'è alcun problema del riferimento, l'uomo e il mondo creano assieme gli oggetti a cui le parole possono poi riferirsi: i riferimenti delle parole, quindi, non sono fatti in senso pieno perché non sono totalmente indipendenti dalla mente.

Come si può salvare il realismo corrispondentista da questi attacchi?

Il realismo accetta che una teoria può risultare falsa indipendentemente dalla grandezza delle sue virtù, quindi bisogna rifiutare l'idea che, per l'indeterminatezza del riferimento, una teoria come quella descritta debba essere vera per forza. L'anello più debole dell'argomentazione sta nel supporre che la teoria ideale rispetti tutti i vincoli che è razionale porre. Tutte le proposizioni della teoria risultano vere, quindi può soddisfare tutti i vincoli che sono esprimibili all'interno della teoria. Tale teoria, però, è espressa in logica del primo ordine, e non c'è motivo per il quale non si possa porre un vincolo esprimibile, per esempio, solo in logica modale, e quindi inesprimibile all'interno della teoria: la teoria potrebbe rispettare oppure no tale vincolo che può essere usato per distinguere il modello inteso dagli alternativi (Lewis 1984).

Vari filosofi hanno reinterpretato l'argomentazione in modo tale da evitare questo tipo di obiezioni, ma più spesso che no tradiscono lo spirito di Putnam. Button (2013) per esempio interpreta l'argomentazione come una *reductio ad absurdum*: se il realista sposa certe teorie sul riferimento, allora, come conclude l'argomentazione, queste portano a sostenere una visione kantiana, che però è inaccettabile, quindi il realismo classico è da rifiutare (ma anche l'antirealismo kantiano lo è, bisogna invece cambiare la teoria del riferimento). Epperò nel momento in cui

Putnam scrive questa argomentazione è chiaramente un pragmatista critico al modo di Rescher, quindi non poteva essere questo il suo pensiero.

Cionondimeno, credo che molte delle intuizioni esposte tramite questa argomentazione siano accettabili a prescindere dall'antirealismo. Putnam stesso, negli anni successivi, è ritornato ad un realismo *tour court*, ma ciò non lo ha mai fatto sentire in dovere di sostenere la legittimità di una sola interpretazione del mondo. L'idea che la verità dipenda dalle categorie mentali individuali è inaccettabile, ma è invece valida e sensata se si sostituisce «verità» con «giustificazione».

Penso vi siano vari modi per accordare l'idea che la realtà vada sempre interpretata, che gli oggetti a cui mi riferiscono siano costruiti applicando schemi mentali alla realtà, con la possibilità delle parole di riferirsi a fatti reali.

Una prima possibilità è quella del *riferimento contestuale*. L'idea è che le parole non abbiano una estensione, ma una *funzione d'estensione*: ad esempio la parola «acqua» ha come estensione una funzione che collega il termine ad enti diversi a seconda del contesto, così che se sono sulla Terra indica l'H₂O, se sono sulla Terra* indica la sostanza XYZ, se sono un cervello in vasca indica una serie di impulsi controllati dal computer, eccetera. Io non so in che contesto sono, ma posso sperare che sia un contesto tale da garantire un certo collegamento (causale, o di altro tipo) con il mondo. Se io ho una intenzione comunicativa devo sperare che il contesto sia tale da garantire il riferimento desiderato, ma affinché io mi riferisca a qualcosa non è necessario che io sappia a cosa mi sto riferendo, è il contesto e non la mia psiche che determina il riferimento. Come posso riferirmi allo spinterogeno *tramite* la conoscenza del meccanico (io non so cos'è, non mi riferisco a niente se non a qualsiasi cosa il meccanico capisca), così accetto come riferimenti delle mie parole qualsiasi cosa il contesto determini come tali (per esempio, posso riferirmi alle esperienze, interpretate tramite i miei schemi, e accettare come riferimento qualsiasi cosa le abbia causate).

Una seconda possibilità è quella di una teoria *controfattuale* del riferimento. Io ora immagino Nessie, il mostro di Loch Ness. A cosa mi sto riferendo, visto che non ho alcuna esperienza di mostri marini? Domani nel lago di Ness viene scoperta una nuova, enorme specie di pesce: è quello il mio riferimento, o deve essere un rettile? Posso dire così: lo è se, lo vedessi (addirittura: vedessi la sua più profonda natura, il pesce noumenico, il pesce-in-sé), sarei disposto a riconoscerlo come l'ente corrispondente alla mia concezione. Questo si collega al problema dell'identità, a quali condizioni sono disposto a riconoscere due enti, in questo caso uno immaginato e uno percepito, come il medesimo ente: di questo parlerò in metafisica.

Ammiro molto questa argomentazione di Putnam, il suo fondersi con la teoria della traduzione di Quine e col *rule-following paradox* del Kripkenstein mi fanno sentire schiacciato come osservassi dal basso un Taj Mahal della filosofia contemporanea, per questo la richiamerò più volte lungo tutto il libro.

Non ci sono prove a favore dell'antirealismo 3: Il Grande Fatto

Nel videogioco Red Dead Redemption 2, un'avventura western ambientata in un enorme mondo aperto, è possibile esplorare una caverna in fondo alla quale vive un personaggio che si presenta come «il diavolo», ma che poi rivela di essere un filosofo. Il contesto ricorda il quadro *A Philosopher by Lamplight* di Joseph Wright. Il personaggio rivela di essere stato emarginato dalla comunità

scientifica per la sua folle tesi sull'inesistenza del linguaggio; e in effetti c'è un grande pensatore del secolo scorso che ha sostenuto davvero (ma scherzosamente) lo stesso, Donald Davidson (2005).

Proprio Davidson (1969) propone lo *slingshot argument par excellence* contro il realismo. Come per il paradosso di Fitch, anche questa argomentazione va fatta risalire a Church (1943), oppure a Gödel (1999-2009 vol. 2). Lungo i decenni è stata formalizzata in un sacco di linguaggi diversi: io propongo una versione formale molto semplificata (basata su Gödel) prima di passare ad una spiegazione ancora più semplificata (basata su Davidson).

Presuppongo che un qualsiasi ente a abbia una proprietà qualsiasi F , che b abbia la proprietà G e che a sia diverso da b . Queste tre proposizioni si scrivono così: Fa , Gb , $a \neq b$ (per esempio, se F è l'essere bagnato e a la pioggia, Fa vorrà dire che la pioggia è bagnata).

Partendo da $a \neq b$ è chiaro che a è definibile come «quella x tale che è uguale ad a e soddisfa $x \neq b$ ». Per esempio, essendo il lampo diverso dal tuono, il lampo è quella cosa diversa dal tuono che è uguale al lampo. Questo si scrive così:

$$\mathbf{P1} \quad a = (x)(x = a \wedge x \neq b)$$

Posso quindi dire che $a \neq b$ equivale ad affermare P1, dire che il lampo è diverso dal tuono equivale a dire che il lampo è quella cosa diversa dal tuono e uguale al lampo:

$$\mathbf{P2} \quad (a \neq b) = (a = (x)(x = a \wedge x \neq b))$$

Posso fare lo stesso ragionamento definendo a partendo da Fa :

$$\mathbf{P3} \quad a = (x)(x = a \wedge Fx)$$

$$\mathbf{P4} \quad Fa = (a = (x)(x = a \wedge Fx))$$

Sostituendo in P1 la definizione di a data in P3 ottengo

$$\mathbf{P5} \quad (x)(x = a \wedge Fx) = (x)(x = a \wedge x \neq b)$$

che potrebbe essere intesa, come esempio, così: quella cosa che è uguale al lampo ed è un fascio di elettroni è uguale a quella cosa che è uguale al lampo e diversa dal tuono. Poiché la parte a sinistra dell'uguale in P5 è la definizione di Fa data in P4, mentre la parte a destra è la definizione di $a \neq b$ data in P2 posso scrivere:

$$\mathbf{P6} \quad Fa = (a \neq b)$$

che posso pensare così: se dico che il fulmine è un fascio di elettroni, dico che il fulmine è diverso dal tuono. Ora posso ripetere tutti questi passaggi dalla prospettiva di b :

$$\mathbf{P7} \quad b = (x)(x = b \wedge a \neq x)$$

$$\mathbf{P8} \quad (a \neq b) = (b = (x)(x = b \wedge a \neq x))$$

$$\mathbf{P9} \quad b = (x)(x = b \wedge Gb)$$

$$\mathbf{P10} \quad Gb = (b = (x)(x = b \wedge Gb))$$

$$\mathbf{P11} \quad (x)(x = b \wedge Gb) = (x)(x = b \wedge a \neq x)$$

$$\mathbf{P12} \quad Fb = (a \neq b)$$

E adesso basta sostituire in P6 la definizione di $a \neq b$ data in P12:

$$\mathbf{C} \quad Fa = Gb$$

Perché una simile conclusione dovrebbe essere problematica per il realismo? Perché, secondo Church e Davidson (ma non secondo Gödel), se due espressioni sono uguali, designano lo stesso fatto. Di conseguenza, se Fa e Gb sono rese vere da dei fatti, essendo uguali, sono rese vere dallo stesso fatto, e dato che Fa e Gb possono essere sostituite due frasi qualsiasi, ogni verità è resa vera dal medesimo fatto. Per rendere più esplicito ciò, Davidson ragiona così:

P1 Immagino che Fa e Gb siano due proposizioni vere qualsiasi, come, «l'erba è verde» e «la neve è bianca»;

P2 per il corrispondentismo, Fa corrisponde al fatto che l'erba è verde, Gb al fatto che la neve è bianca;

P3 posso ora comporre una proposizione A così: $(x \wedge Fa) = x$ (Davidson la legge come «(Diogene, e l'erba è verde)=(Diogene)», quale sia il significato preciso di «Diogene» è irrilevante)

P4 a logica, secondo Davidson, $(x \wedge Fa) = x$ equivale alla sola Fa , e quindi per presupposto x ed Fa devono indicare lo stesso fatto;

P5 scrivo ora un'altra proposizione che nomino B : $(x \wedge Gb) = x$; quindi anche Gb e x devono indicare lo stesso fatto;

P6 A e B sono entrambe uguali a x , quindi saranno anche uguali tra di loro: $(x \wedge Fa) = (x \wedge Gb)$, e poiché sono rispettivamente uguale a Fa e Gb posso scrivere $Fa = Gb$;

C quindi Fa e Gb , «l'erba è verde» e «la neve è bianca», devono corrispondere allo stesso fatto, e ciò si può ripetere per qualsiasi coppia di frasi vere. Quindi c'è un unico fatto: la verità!

Nonostante la loro poeticità, queste conclusioni sarebbero distruttive: sarebbe un po' come dire che l'unica cosa a cui una proposizione corrisponde è il valore di verità medesimo, un'idea alla quale aveva già accennato Frege (2007). Ciò può essere interpretato in due modi. In primo luogo può costituire una *reductio ad absurdum*: se il corrispondentismo è vero esiste un unico fatto, o comunque ogni proposizione è vera in virtù di uno stesso fatto, e questo è palesemente falso. In secondo luogo può invece indicare che, poiché ogni frase si riferisce al valore di verità medesimo, ciò che la rende vera non è un fatto in senso proprio. In entrambi i casi, il realismo corrispondentista ne esce malmenato. Ma l'argomentazione ha dei punti deboli.

Il problema più discusso riguarda la frase $(x \wedge Fa) = x$. Cosa vuol dire il simbolo « \Rightarrow »? L'equivalenza *materiale* è una relazione del tipo che può apparire in una domanda o richiesta: per esempio posso chiedere «dimostra che $2+2$ è *materialmente equivalente* a 4 ». Se dimostro che, in questo caso, lo sono, cioè che sono entrambi veri o entrambi falsi, dico che sono *uguali*, e questo è il senso del simbolo (potrò dire che sono anche logicamente equivalenti se dimostro che è impossibile che abbiano valori diversi, e quindi che l'uguaglianza deve essere sempre vera). Ma il

fatto che le due espressioni siano equivalenti non è sufficiente ad affermare che si riferiscono allo stesso fatto: «è un triangolo» ed «è un trilatero» sono logicamente equivalenti, ma uno è reso vero da un fatto sugli angoli e l'altro da un fatto sui lati (cfr. Barwise & Perry 1981, Taylor 1985).

Perché in $(x^{\wedge}Fa) = x$ ed Fa dovrebbero essere resi veri dagli stessi fatti? Qualsiasi sia il fatto a cui corrisponde Fa , anche x deve corrispondervi, e questo vale per qualsiasi proposizione io metta al posto di Fa . Pare quindi che x si riferisca semplicemente a tutti i fatti, e quindi inglobi i fatti a cui si riferiscono Fa e Gb : è un po' come se $(x^{\wedge}Fa)$ e $(x^{\wedge}Gb)$ significassero «tutte le parti della macchina, e il motore» e «tutte le parti della macchina, e le ruote», e in tal caso è vero che entrambe le congiunzioni equivalgono al solo «tutte le parti della macchina», ma motore e ruote continuano a riferirsi a fatti diversi (Stoutland 2003 specifica che in $(x^{\wedge}Fa)$ i congiunti devono indicare lo stesso oggetto se x ne indica uno solo, ma non vedo perché dovrebbe darsi questo caso).

Questo evidenzia, infine, il problema dell'interscambiabilità (come in Frege 2007): è vero che «Superman» e «Clark Kent» si riferiscono allo stesso fatto, ma in una frase come «Lois Lane non sa che Clark Kent è Superman» non li posso scambiare per ottenere «Lois Lane non sa che Superman è Superman», la prima frase è vera e la seconda è falsa. Similmente, anche accettando che $(x^{\wedge}Fa)$ indichi lo stesso fatto di Fa , questo non significa che da $(x^{\wedge}Fa) = Gb$ io posso formare $Fa = Gb$.

Il pluralismo non è un'alternativa al realismo

Le filosofie-tipo esposte finora non sono divise in modo ferreo, ma nessuno pare essere realista (o antirealista) verso *tutti* i campi del discorso: tutti credono che la scienza faccia affermazioni vere e tutti credono che sia vero che la fenice è immortale, ma mentre tutti credono che la verità della scienza dipenda dai fatti, non tutti lo credono per la fenice. Mi chiedo allora: perché uno non può affermare di sostenere più di una teoria della verità?

È ovvio che io non posso sostenere contemporaneamente che è vero ciò che corrisponde ai fatti *ed* è vero ciò che è coerente. Se io credessi ciò, e si accetta che l'esistenza dello Yeti non è un fatto ma è logicamente coerente, allora dovrei dire lo Yeti esiste (per coerenza) e non esiste (per corrispondenza). Potrei dire di accettare due teorie contemporaneamente solo se sono logicamente equivalenti, cioè se le due teorie assegnano esattamente gli stessi valori di verità a tutte le proposizioni; in caso contrario, sarebbe *relativismo*. Epperò nulla mi vieta di sostenere diverse teorie non-equivalenti *per* diversi campi del discorso: non posso sostenere più teorie della verità, ma posso sostenere che la verità è fattuale per certi campi e non-fattuale per certi altri.

Questa è la base della *teoria pluralista* della verità. L'idea è quella di dividere le affermazioni possibili in classi per individuare cosa *rende vere* le affermazioni di una certa classe, cioè quali sono i *truthmaker*: si può dire, per esempio, che le proposizioni scientifiche sono rese vere dalla corrispondenza ai fatti naturali, quelle matematiche dalla coerenza, quelle etiche da leggi o convenzioni sociali, quelle religiose dalla loro utilità, eccetera (cfr. Wright 1992). Non è relativismo, non si dice che ciò che è vero per uno può non esserlo per un altro a seconda della prospettiva, a seconda della teoria scelta.

È però molto facile confondere pluralismo e relativismo: ho avuto grandissimi problemi a interpretare correttamente la teoria pluralista proprio perché continuavo a vederla come ciò che

non era. Personalmente, per eliminare questi fraintendimenti, questa confusione tra pluralismo e relativismo, ho trovato utile pensare alla funzione *metalinguistica* del linguaggio.

Un *metalinguaggio* è un linguaggio che parla di un altro linguaggio. Quando dico «*the sun will rise tomorrow* significa che il sole sorgerà domani» uso l'italiano come metalinguaggio per parlare dell'inglese. Avrei potuto dire anche così: «*the sun will rise tomorrow* se e solo se il sole sorgerà domani». Il «se e solo se» (abbreviabile in un comodo «sse») indica qui l'equivalenza, dice che una parte è necessaria e sufficiente all'altra, che la parte in inglese è vera se lo è quella in italiano e viceversa. Ora posso intendere la parte a sinistra del sse come relativa ad una proposizione, e quella a destra del sse come relativa al fatto che la rende vera: la proposizione è vera se si dà il fatto di cui parla. Quindi:

- x è vero sse corrisponde ai fatti.

« x » è una proposizione in un qualche linguaggio, «...è vero...» è il predicato in metalinguaggio, e «...corrisponde ai fatti» è il suo equivalente. Ora, prendendo ispirazione da Moore 1903, voglio chiedermi se tale definizione di *verità* sia a sua volta vera. Qualsiasi risposta a tale domanda dovrà avere la forma:

- « x è vero sse corrisponde ai fatti» è vero sse y .

Bisogna fare attenzione, perché se si interpreta la frase al valore nominale (com'è facile) si finisce per dire delle assurdità. Sembra, infatti, che io stia affermando che una certa definizione di «verità» è vera solo se corrisponde ad un certo altro criterio di verità (e si può andare avanti così all'infinito). Questo porta ad un caso come quello dello Yeti.

Il punto è che la frase tra virgolette conta come linguaggio rispetto al successivo «...è vero...», quindi i due predicati di verità non possono essere allo stesso livello: la « x » deve essere linguaggio, il primo predicato metalinguaggio, e il secondo meta-metalinguaggio. Perché bisogna interpretare così?

Dopo Gödel, il maggior logico del secolo scorso è Alfred Tarski, della gloriosa scuola polacca. La sua fama deriva soprattutto dalla sua analisi del concetto di verità, e infatti la definizione che ho proposto ha la stessa forma del suo *convenzione-T*, « x è vero sse x » (Tarski 1983). Tarski nota che molti dei paradossi legati al concetto di verità originano quando il linguaggio contiene il suo proprio predicato di verità, cioè quando il linguaggio può affermare «è vero» di sue frasi. C'è allora un modo diretto e semplice per risolvere il problema: il predicato di verità deve comparire solo ad un livello superiore, solo in metalinguaggio, e non allo stesso livello del linguaggio della proposizione a cui si riferisce.

Tarski non pensava a questa teoria come estendibile all'analisi della lingua naturale: la sua è una filosofia della logica, non del linguaggio ordinario; la lingua naturale può creare frasi paradossali come «questa frase non è vera» perché può porre il predicato di verità allo stesso livello degli altri elementi, mentre nelle *lingue ideali* ciò è inaccettabile. Quella che lui presenta come verità è quindi la verità dei linguaggi formali, non è necessariamente la verità della lingua naturale; ma visto che la lingua naturale esprime proposizioni, e che almeno alcune di queste sono esprimibili anche in linguaggio formale, ciò non ha impedito ai filosofi di usare il lavoro di Tarski per parlare della verità in senso generale.

Orduunque, riscrivo la frase al modo di Quine (2004), evidenziando coi pedici i due livelli di metalinguaggio:

- « x è vero₁ sse corrisponde ai fatti» è vero₂ sse y .

Ora devo chiedermi se «vero₁» e «vero₂» esprimano lo stesso significato, o non siano invece due parole diverse, due tipi di verità, come un *vero* e un *vero**. La teoria pluralista, di per sé, può sembrare non impedire ciò: se diversi tipi di frase hanno diverse condizioni di verità, può ben essere che la definizione di verità abbia un truthmaker diverso da tutti gli altri. Suppongo sia così, e scrivo:

- « x è vero₁ sse corrisponde ai fatti» è vero₂ sse è coerente.

Questo potrebbe essere problematico: la teoria della verità non basta più alla filosofia, serve una *teoria della verità*₁ e una *teoria della verità*₂! I pluralisti, come accennato, non sostengono una cosa del genere, non sostengono che «verità» sia un termine ambiguo che esprime più significati in diversi contesti (Wright 1996). Il pluralismo è una teoria della verità e punto, senza pedici: accetta una verità *multipla*, cioè che vi siano più criteri di verità, ma questi criteri devono essere tutti allo stesso livello, non a più livelli di metalinguaggio, cioè devono far parte di un'unica definizione di verità e non creare più definizioni. Non c'è una verità fattuale e una non-fattuale, bensì il significato di «verità» ingloba aspetti sia fattuali che non-fattuali.

Volendo, anche una teoria *monista* può essere riscritta in stile pluralista. Per esempio, un bambino mi chiede cosa voglia dire «colorare bene» una figura, io dico « x è colorato bene sse è del suo colore naturale», ma potrei anche specificare: « x è colorato bene sse: se è un mare, sse è blu; se è un albero, sse è verde...». Quindi, un esempio di verità pluralista potrebbe essere:

- x è vero sse [(se x è una proposizione analitica, sse il concetto del predicato è incluso nel concetto del soggetto) o (se x è una proposizione sintetica, sse corrisponde ai fatti)].

Anche questa è una frase che ho spesso frainteso. Pensavo che la teoria pluralista della verità, per non rischiare di autorifiutarsi, dovesse includere un truthmaker per la definizione stessa; al che, supponendo che la definizione sia di tipo sintetico (perché deriva dall'analisi dell'uso delle parole), scrivevo:

- « x è vero sse [...]» è vero sse corrisponde ai fatti.

E così si finisce di nuovo in un caso paradossale come quello dello Yeti, oppure si è costretti ad accettare l'esistenza di due tipi di verità indipendenti. Ma è errato, perché per il pluralismo la definizione di verità è una, è un'unità. Dovevo scrivere così:

- « x è vero sse è positivo» è vero sse è positivo

dove è positivo (ad esempio) ciò che, se è analitico, ha il predicato incluso nel soggetto, e, se è sintetico, corrisponde ai fatti. Cioè: ogni proposizione è vera se è positiva, ma come si possa controllare se una proposizione sia positiva dipende dal suo tipo. Anche se l'idea del pluralismo fa subito pensare a paradossi logici, una teoria come questa è invece coerente e di buon senso.

Ovviamente una definizione plurima di verità dovrebbe essere molto più complessa di così. Prendo per esempio « $2+2=4$ » e «questo tavolo è di legno». Sono chiaramente proposizioni vere in modo diverso. Cosa dire della proposizione « $2+2=4$ e questo tavolo è di legno»? È vera se i

congiunti sono veri; quindi una frase della forma «x e y» è sempre vera nello stesso modo, contrariamente al solo «x». Posso però creare infinite proposizioni composte, infinite e complessissime combinazioni di segni, e ognuna avrebbe una condizione a parte: quindi è impossibile esplicitare la condizione di verità di ognuna, ci si deve accontentare di un metodo per calcolare i valori di verità di volta in volta (cfr. Tappolet 2000).

C'è però un problema: quali e quanti diversi tipi di proposizione esistono? Servirebbe una oggettiva *tassonomia della verità*. Il pericolo è che si arrivi a produrre una classificazione a grana così fine delle proposizioni, ognuna con un suo criterio di verità così specifico, che, alla fine, ogni proposizione risulti vera *in un modo o nell'altro* (cfr. Pedersen 2006). Per esempio, perché dovrebbe essere impossibile ritenere «la terra è piatta» un tipo di frase *sui generis*, distinta dagli altri enunciati cosmologici, fattualmente falsa ma culturalmente vera? Il pluralismo, di per sé, non prevede nulla che impedisca un'operazione del genere.

A questo punto sorge un altro sospetto: che utilità può avere il definire la verità in questo modo? Posso accettare, ma solo per fare l'esempio, che la verità sia equivalente alla positività. Fatto sta che vi saranno proposizioni *positive e fattuali* (vere in virtù di fatti) contro altre *positive e non-fattuali* (vere a prescindere dai fatti) o *non-positive* (false): anche se esistono proposizioni che sono vere in modo non-fattuale, perché dovrebbero interessarmi? Perché invece non posso dire che una è *vera e positiva* e l'altra *positiva anche se falsa*? Perché non dovrei sviluppare una *teoria della fattualità*, considerandola una verità di qualità superiore?

In effetti ci sono dei buoni motivi: nelle lingue naturali «vero» è usato in modo apparentemente non-fattuale. Tutti gli uomini credono che «uccidere è male» e « $2+2=4$ » siano entrambe frasi vere, ma non si capisce come sia possibile che lo siano entrambe allo stesso modo (cfr. Tappolet 1997, Pedersen 2010). Tutti gli uomini credono che «uccidere è male» sia vero, lo crede addirittura chi non crede all'esistenza di fatti etici: e questa è una splendida manifestazione di antirealismo. Ma è così? Forse la società occidentale parte dal presupposto inconscio, foss'anche errato, che i fatti etici esistono, e coloro che non vi credono si adeguano all'uso della lingua. Se un giorno si scoprissero dei fatti etici, se per assurdo si scoprisse che il bene è identico a qualche caratteristica delle particelle, cosa direbbero coloro che non credono ai fatti etici? Forse «credevamo che il male fosse non-fattuale, e invece era fattuale», forse addirittura «*davvero vero!*». In ogni caso, sembra che la verità fattuale abbia un primato rispetto a ciò che, pur potendo essere riconosciuto vero, non è fattuale: fattualità e non-fattualità possono essere entrambi aspetti della positività o qualche altro criterio, ma devono avere una gerarchia.

Il deflazionismo non è un'alternativa al realismo

Il *deflazionismo* non è solo una teoria della verità, ma un progetto filosofico generale. Il nome indica uno sgonfiamento: ci sono problemi filosofici percepiti come complessi e importanti, il deflazionismo li sfiora e li sgonfia rivelandoli come semplici e banali. La filosofia di Wittgenstein (2014 soprattutto) può essere considerata un deflazionismo globale (anche detto *quietismo* o *approccio terapeutico*), il suo scopo non è risolvere i problemi filosofici ma mostrare perché sono cosette semplici e poco importanti delle quali non preoccuparsi.

Russell, come Frege prima di lui e molti altri dopo, ha sostenuto un deflazionismo nei confronti del concetto di *esistenza*. Posso immaginare una *mela rossa*, ma ha senso parlare di una *mela rossa esistente*? Se io tolgo ad una *mela rossa esistente* la proprietà dell'esistenza, cosa le tolgo esattamente?

Se dico che una cosa è una mela, o che è rossa, sto già dicendo che è esistente; quindi dire che una certa cosa è una *mela esistente* è proprio identico al dire che è una *mela* e punto. Infatti: se due oggetti condividono una proprietà condividono anche una certa spiegazione in merito alla tale proprietà, due oggetti che condividono l'*essere rosso* condividono anche l'averne una superficie tale da riflettere certe lunghezze d'onda e non altre; ma io posso dire che due oggetti sono esistenti anche se non condividono nulla, la mia esistenza ha una spiegazione e la casa ne ha un'altra (Russell 2005).

La teoria di Russell è controversa. Emanuel Rutten, un filosofo olandese, ha presentato una ricerca nella quale ritiene di dimostrare che, se l'esistenza non è una proprietà, allora non esistono proprietà universali, e così, se si accetta che l'essere naturale è una proprietà, allora deve poter esistere anche qualcosa di non-naturale (Rutten 2014). Ovviamente un naturalista non può accettare queste conclusioni; io ho provato a rifiutarle ma non ci sono riuscito. Ciò nonostante, la teoria di Russell rimane rilevante poiché il suo ragionamento è riproponibile quasi uguale in epistemologia.

Tutte le teorie della verità che ho visto finora sono *sostanziali*, cioè considerano la verità come una proprietà posseduta dalle proposizioni, un *modo d'essere* delle proposizioni. Le teorie deflazionarie negano ciò: non c'è alcuna definizione particolare di verità perché, come l'esistenza, neanche la verità è una proprietà, dire che qualcosa è vero anziché falso non dà alcuna informazione rilevante.

Frank Ramsey è stato davvero il Mozart del pensiero: ha dato contributi alla matematica, alla filosofia e all'economia che sono riconosciuti come tra i più originali e innovativi di tutto il Novecento prima di morire all'età di 26 anni (Misak 2020). Il suo maggior risultato in filosofia è stato quello di aver sviluppato, sulla base di Frege (2007, 2019), la prima teoria deflazionaria della verità: se io dico che «la neve è bianca» non dico nulla di meno rispetto a «è vero che la neve è bianca» (né a «è vero che è vero che la neve è bianca», e via così). La verità è *ridondante* perché ogni proposizione, per il solo fatto di essere tale, afferma la sua propria verità. L'errore dei filosofi sta nel credere che la verità sia un qualcosa da *scoprire* come se la parola si riferisse a qualche oggetto, quando invece è una questione grammaticale, riguarda la forma delle frasi (Ramsey 2020).

Ma allora perché si parla tanto della verità, perché la lingua prevede una parola per esplicitarla? Ecco in questo punto tutta la teoria deflazionista: una volta espressa la funzione di «verità», il motivo per cui si usa la tal parola, si è già detto tutto ciò che c'è di rilevante sul tema.

Si può dire, per esempio, che il predicato «è vero» esprime l'atteggiamento del parlante, il proprio sottoscrivere ciò che si sta dicendo, o comunque che sia un modo di dare enfasi (Strawson 1949). Un'altra spiegazione proposta da Ramsey è che, come «io» è un pronome che sostituisce il mio nome, così «è vero» dovrebbe funzionare come una *pro-frase* che sostituisce altre espressioni; per esempio:

- «Giovanni crede che *sta piovendo*, e sta piovendo» diventa: «crede che *sta piovendo*, ed è vero»;
- «Giovanni, se dice x allora x , se dice y allora y , se dice z allora z ...» diventa: «Giovanni dice solo la verità»;

e questo è tutto ciò che giustifica l'esistenza di questa strana parola, ed è anche ciò che crea confusione poiché è la mancanza di *pro-frasi* in lingua naturale che mi fa credere che la verità sia una entità da scoprire (cfr. Grover et al. 1975).

Nella *Lettera a Tito*, San Paolo (o un suo epigono più probabilmente) cita Epimenide, filosofo cretese (ma lui lo chiama *profeta*) e gli attribuisce questa frase: «i cretesi sono sempre bugiardi,

brutte bestie e fannulloni» (Tt 1, 12). Un cretese che dice che tutti i cretesi mentono... Ovvio: se tutti i cretesi mentono anche lui deve mentire, ma allora la frase è vera e non sta mentendo. È un po' come dire «questa frase è falsa»: se è falsa è vera, e se è vera è falsa. Trovare il modo di gestire frasi del genere è una delle difficoltà della teoria deflazionaria.

Poiché nel deflazionismo « x è vero» è equivalente al solo « x » posso scrivere:

- x è vero sse x .

che è la convenzione-T di Tarski, e che per i deflazionisti costituisce tutto ciò che è necessario dire e sapere sulla verità. Allora «questa frase è falsa» diventa:

- «questa frase è falsa» è vero sse questa frase è falsa.

Il *paradosso del mentitore* è uno di quelli dovuti al fatto che la lingua naturale contiene il suo proprio predicato di verità: è una frase che fa una affermazione sul suo stesso valore. Questo paradosso è un problema per tutte le teorie della verità, ma per il deflazionismo in modo particolare perché per rispettare la teoria non posso aggiungere clausole particolari o dare più elaborate definizioni di «verità» che vietino la formazione di frasi di questo tipo.

Ma c'è una soluzione molto semplice e compatibile con il deflazionismo. Per evidenziarla, scambio la posizione delle due frasi:

- questa frase è falsa sse è vero che «questa frase è falsa».

Cosa vuol dire? Per il deflazionismo ogni proposizione presuppone la sua verità, dire « x » è come dire « x , ed è vero che x »; quindi «questa frase è falsa» diventa «questa frase è falsa, e questa frase è vera». Ma questa è una pura e semplice contraddizione, le contraddizioni sono sempre false, le due frasi dello schema sono equivalenti quindi sono entrambe false. E quindi non c'è alcun paradosso, «questa frase è falsa» è falsa e punto.

Questa soluzione, proposta dal grande logico Prior, può essere fatta risalire a Peirce se non addirittura a Buridano, il filosofo medievale famoso per il suo asino (Prior 1976). Ovviamente è, come tutte le soluzioni, controversa, ma è stata difesa da filosofi come Eugene Mills (1998; si veda anche Blackmon 2005).

Un altro problema riguarda il capire se il deflazionismo sia una teoria sulle frasi o sulle proposizioni. Io ora scrivo:

- «la neige est blanche» è vero sse la neve è bianca.

Se accetto l'espressione tra virgolette come una frase in lingua, ciò che ho detto può essere necessario, ma non sufficiente, alla verità: affinché «la neige est blanche» sia vero è necessario che la neve sia bianca solo se si dà anche il caso che «neige» voglia dire *neve*, «blanche» *bianca*, eccetera.

E se la intendo come teoria sulle proposizioni? Scrivo:

- *la neve è bianca* è vero sse la neve è bianca.

E questo è assurdamente poco interessante: mi sto limitando ad evidenziare l'equivalenza tra due proposizioni, sto dicendo che « x » ha gli stessi valori di verità di «è vero che x », ma non sto dicendo

nulla sul modo in cui dovrei calcolare il valore di verità, non si capisce in quali casi la proposizione deve essere considerata falsa e in quali vera. E che teoria della verità è una che non dice come assegnare valori di verità? Se la teoria accetta le proposizioni come *truthbearer* (ciò che può essere vero o falso) perché non mi dice anche quali sono i truthmaker? (Jackson et al. 1994)

Questo è il maggior problema del deflazionismo: la convenzione-T di Tarski è accettato da molte altre teorie (Tarski stesso credeva che la sua teoria fosse di tipo corrispondentista) ma solo la deflazionaria *si limita* allo schema senza aggiungere null'altro, fino al punto da sembrare incompleta: Ramsey fatto il quadro e non ha messo la firma, come diceva Altafini, anche se non è colpa sua essendo morto prima di poter completare il libro che stava programmando. Al che, alcuni filosofi interpretano il deflazionismo non come una teoria autonoma, ma una teoria da fondere alle altre: per esempio, Quine sosteneva il deflazionismo e il coerentismo contemporaneamente, mentre oggi molti credono che il deflazionismo, per poter essere accettato, debba dimostrarsi compatibile con il corrispondentismo.

Ma quello esposto è solo *un* deflazionismo. Ho detto che la teoria può riguardare frasi o proposizioni, ma c'è anche chi dice che riguarda frasi *interpretate*, in cui il significato dei termini è già assegnato. Cosa dire del «sse»? Io l'ho interpretato come una equivalenza materiale, ma c'è chi lo interpreta invece come una equivalenza logica necessariamente vera, e c'è addirittura chi ritiene debba indicare che le due parti hanno lo stesso significato, una richiesta ancora più forte, cioè che pone più vincoli, di quella dell'equivalenza logica. Infine, c'è chi nega direttamente la convenzione-T favorendo una definizione esplicita della verità deflazionaria. Quante teorie deflazionarie esistono? Ne ho contate dodici varietà (Stoljar & Damnjanovic 2014), e sono in aumento: è ben possibile che tra queste ve ne siano di compatibili con il corrispondentismo.

Detto tra parentesi: naturalismo e antinaturalismo

Scrivo delle affermazioni:

1. «uccidere è male»;
2. « $2+2=4$ »;
3. «*io* ho mangiato»;
4. «ieri ha piovuto, domani pioverà»;
5. «Sherlock Holmes è un detective, è un'icona popolare, ed è più intelligente di qualsiasi detective reale»;
6. «se avessi studiato di più, ora sarei laureato».

La prima rappresenta le proposizioni etiche. A cosa mi riferisco quando dico che «uccidere è male»? Quale fatto rende vera la tale frase?

La seconda rappresenta le proposizioni matematiche. A cosa si riferisce? Il «due», seguendo Frege 2019, può essere inteso come l'insieme di tutte le coppie, tale è la sua estensione; ma con calcoli più complessi e astratti diventa difficile rispondere. Qual è l'estensione di «infinito»? Dove sono, nell'universo, i fatti che rendono vere tali frasi?

La terza evidenzia un problema con l'identità. «Io» è un *indexicale*, il suo riferimento cambia col contesto, cambia a seconda di chi è che dice «io». Se è vero che «io, un'ora fa, ho mangiato» deve

esiste un qualche fatto tale che l'io che parla adesso è lo stesso che ha mangiato un'ora fa. Quale fatto può fare ciò?

La quarta rappresenta le proposizioni temporali. Un fatto è qualcosa di reale, qualcosa che esiste; com'è possibile, allora, che una affermazione sul passato o sul futuro, cioè su ciò che è esistito o esisterà, corrisponda ad un fatto che esiste?

La quinta evidenzia un problema con le entità fittizie. «Sherlock Holmes è un detective» a che fatto si riferisce? Posso pensare ad un fatto letterario, come se la frase dicesse che lo è *nei romanzi*, che è *descritto come* un detective. Però questo non può dirsi per «Sherlock Holmes è un'icona popolare». Potrebbe riferirsi ad un fatto sociale allora. Ma l'ultima, «Sherlock Holmes è più intelligente di qualsiasi detective reale», non è riconducibile a nessuno dei due casi, mescola finzione e realtà: quale fatto può renderla vera?

La sesta riguarda i controfattuali. Si può reinterpretare l'ultima frase su Sherlock Holmes così: «*se* Sherlock Holmes esistesse, *sarebbe* più intelligente di ogni altro detective». Ma un controfattuale come questo parte assumendo vero ciò che *non* è un fatto, si chiama così perché è un tipo di affermazione che va contro i fatti: dunque come può essere vera, cioè corrispondervi? Forse è un caso di *ex falso quodlibet*: poiché una frase della forma «*se x allora y*» è falsa solo se *x* è vero e *y* è falso, ad un presupposto *x* falso si può far seguire qualsiasi *y* e la frase sarà sempre e comunque vera. Ma, fosse così, i controfattuali sarebbero sempre veri, e invece «*avessi studiato di più, ora Mosca sarebbe la capitale del Messico*» pare proprio falsa: cosa differenzia tale frase dalle altre?

Le questioni sollevate dalla prima frase le discuterò in etica; quelle sollevate dalla seconda, terza, e quarta in metafisica, nello specifico in filosofia della matematica, dell'identità, e dello spaziotempo; quelle sollevate dalla quinta e sesta non le affronto. Ora voglio parlarne solo da un punto di vista generale, ovvero mi chiedo: che tipi di fatti ci sono?

Tutte le frasi viste paiono poter essere vere; epperò non è sicuro che possano esserlo se si accetta il corrispondentismo e gli unici fatti sono quelli descritti dalle scienze naturali, come particelle e loro insiemi: le proposizioni etiche potrebbero richiedere fatti etici, quelle matematiche fatti matematici, vari tipi di fatti *sui generis* non riducibili a quelli naturali, a quelli previsti dalla descrizione scientifica del mondo. Coloro che accettano l'esistenza di soli fatti naturali sono chiamati *naturalisti*. Coloro che accettano l'esistenza reale anche di altri fatti (che siano insiemi, valori, essenze, enti non-esistenti, angeli, mostri...) devono quindi essere *non-naturalisti*. Mettendo insieme teoria della verità e posizione in merito ai fatti individuo quattro filosofie:

- naturalismo realista;
- naturalismo antirealista;
- non-naturalismo realista;
- non-naturalismo antirealista.

Il *naturalismo realista* è una filosofia che si associa alla visione scientifica del mondo: la verità è corrispondenza ai fatti, e non esistono che fatti naturali. Ad intuito, pare un'idea oggi molto comune; se però si cerca di nominare uno specifico filosofo che senza dubbio aderisce a tale posizione, be', non è facile individuarlo.

Un naturalista crede nell'esistenza della sola natura, ma cos'è la natura? Se è ciò che la scienza descrive *ora*, è una posizione sicuramente falsa (poiché la scienza è in evoluzione), nessuno sarebbe naturalista. Se è ciò che la scienza *descriverà*, non si può dire cosa è e cosa non è natura, anche Dio

potrebbe esserlo, essere naturalista sarebbe un atto di fede. Se il naturalismo è, per esclusione, *non-supernaturalismo*, cioè ateismo, allora un terrapiattista potrebbe essere un naturalista, è una definizione troppo vaga. Forse questo spiega il successo del naturalismo, ognuno lo interpreta come meglio crede; ma proprio per questo il naturalista di uno è il non-naturalista dell'altro, e viceversa.

Nagel, filosofo ateo di larghe vedute e dotato di pensiero profondo in più settori, si definisce naturalista, ma rifiuta l'idea che la mente sia riducibile ad enti fisici o che possa essere spiegata dall'evoluzione biologica, e non è chiaro in che altro senso possa essere naturale (Nagel 2015). Russell pure si descriveva come un naturalista, ma è sempre rimasto influenzato dall'idealismo, non pensava che la fisica offrisse una descrizione completa della realtà, bensì credeva nell'esistenza di proprietà intrinseche diverse dalle fisiche ma reali: neanche lui pare un naturalista in senso pieno (Russell 2013). Forse gli unici ad essere indubabilmente naturalisti fino al midollo sono i Churchland, Paul e Patricia, marito e moglie uniti nel sacro vincolo del nichilismo. (e.g. Churchland 1992).

Il problema di un naturalista realista purosangue, uno che crede solo in enti fisici, è che non può rendere merito dell'intuitiva verità di frasi che cercano di riferirsi a fatti chiaramente astratti. Ha solo due possibilità: o segue una *error theory*, cioè sostiene che quelle frasi esprimono classi di affermazioni sempre false, o segue un *non-cognitivism*, cioè sostiene che quelle frasi non esprimono proposizioni perché non vogliono essere affermazioni di fatti e quindi non hanno un valore di verità. Anche se *error theory* e *non-cognitivism* sono accettati da chiunque per almeno alcuni settori della discussione, credere che possano valere per tutti i casi elencati è indesiderabile e contrastante con il fatto evidente che i parlanti cercano davvero di fare affermazioni su quei temi.

Il *naturalismo antirealista* riconosce l'esistenza dei soli fatti naturali, ma non crede che la verità stia nel corrispondere a tali fatti.

Il naturalismo realista può sembrare una posizione scienziata, ma i filosofi che con più durezza e sicurezza hanno affermato il primato della scienza sono stati antirealisti. I positivisti logici ritenevano sensate, quindi o vere o false, solo le proposizioni verificabili, cioè solo quelle dimostrabili tramite un test empirico (se sintetiche) o analisi logica (se analitiche). Ma il *principio di verifica* ha senso solo se la verità non sta nel corrispondere ai fatti: se uno accetta il corrispondentismo, allora una certa affermazione potrebbe ben essere inverificabile e comunque corrispondere ai fatti e quindi essere vera e sensata. Il concetto di verità positivista non è fattuale, ma costruito, si costruisce la verità acquisendo la possibilità di verificare le affermazioni (cfr. Schlick 1936).

È sorprendente che filosofi così fedeli alla scienza sostenessero un ideale di verità non-fattuale, e infatti, secondo Dummett 2001, i positivisti logici non erano sempre coscienti di questa implicazione del loro lavoro. Lo erano invece i loro più grandi figli, Quine e Putnam, soprattutto quest'ultimo che ha sostenuto esplicitamente sia il naturalismo che forme di antirealismo (Putnam 1993). Se si accetta il naturalismo come non-supernaturalismo allora anche quasi tutti i filosofi continentali contemporanei possono essere naturalisti antirealisti. Per costoro le frasi espresse sopra non costituiscono un problema: possono essere vere anche se non corrispondono a fatti.

Il *non-naturalismo realista* può essere considerata la posizione standard di buona fetta della filosofia medievale, nonché di razionalisti come Leibniz, *idealisti oggettivi* come Schelling (1988), ed eredi dell'idealismo come Whitehead (anche se, a causa delle potenti critiche portate avanti da Russell e

Moore, ben pochi oggi si direbbero idealisti). È la soluzione più diretta e coraggiosa: le frasi esposte sono vere e lo sono in modo fattuale in quanto esistono anche fatti non-naturali.

Ma questo coraggio deve bastare anche per affrontare a viso aperto domande molto difficili: com'è possibile conoscere fatti non-naturali? L'idea che esistano fatti in grado di dare un valore di verità ad *ogni* tipo di frase, anche quelle relative ad enti fittizi, è estrema, richiama l'ontologia barocca di Meinong. L'alternativa è accettare l'error theory e il non-cognitivism per alcuni settori: ma perché non fare il passo successivo, e accettarle per tutti? O perché non farlo nell'altra direzione, e accettare che la verità di cui si parla non è di tipo fattuale?

Il *non-naturalismo antirealista* può accettare l'esistenza di fatti non-naturali, o anche negare l'esistenza di fatti naturali, ciò non è determinante perché la sua verità non è fattuale. Si trovano qui gli idealisti più radicali come Hegel, ma anche scienziati come Peirce e il chimico Michael Polanyi (e.g. 2012 e 2018). Discutibile il caso di Dummett: in quanto cattolico difficilmente lo si definirebbe naturalista, ma il suo antirealismo è solo parziale, è stato lui che lo ha riportato in auge ma credeva anche fosse autocontraddittorio applicarlo a tutti i campi del discorso.

LETTURE PER APPROFONDIRE

L'argomentazione dei modelli di Putnam si appoggia a Gödel e Cantor: per approfondire in modo divulgativo questi personaggi si vedano Odifreddi 2018, Barrow 2006 e Aczel 2015.

Per altre forme di antirealismo, consiglio Fleck 1981, Bonjour 2002 e Rescher 1982.

Al contrario, tra i realisti naturalisti, si veda Papineau (1993), filosofo inglese ma nato a Como. In generale, tutti i discepoli del nuovo ateismo difendono un naturalismo superficiale (e.g. Dawkins 2007, Hitchens 2007, Harris 2012, Odifreddi 2011), anche se ad analisi più approfondita, come nel caso di Russell, raramente si può parlare di un naturalismo purosangue.

Il deflazionismo va integrato con le teorie di Horwich 1998, 2004 e 2010.

Epistemologia: teoria della giustificazione

Secondo ragionamento: un fondazionalismo aperto

In questo secondo capitolo, a differenza del primo, oltre al valutare le teorie classiche, propongo anche una tesi personale sulle proprietà della conoscenza. Nello specifico, il ragionamento che guida il capitolo è il seguente:

- P1** la conoscenza deve essere indefettata;
- P2** la conoscenza deve essere aperta;
- P3** la coerenza di giustificatori indipendenti non è truth-conducive;
- P4** ma il coerentismo lo nega;
- P5** e il fondazionalismo nega l'apertura epistemica;
- C** quindi nessuna delle due teorie è soddisfacente.

La conoscenza deve essere indefettata 1: i casi di Gettier

La verità *in senso pieno*, la fattualità, è del tutto indipendente da me: io non posso conoscere i fatti in modo diretto. Nella teoria della giustificazione non devo avere la pretesa di elencare fatti: devo invece capire quand'è che posso legittimamente chiamare conoscenza una credenza, cioè quando sono legittimato a crederla vera (anche se potrebbe non esserlo). Prima di proseguire, però, devo accertarmi che l'analisi della conoscenza proposta sia corretta, e quindi che la giustificazione sia realmente una componente di tale concetto, o che non vi siano altre componenti essenziali che ho tralasciato di discutere.

Ho visto un video su Facebook. Delle persone vengono chiamate per un colloquio di lavoro, entrano nell'ufficio del datore, che ha alle spalle ad una grande finestra con veduta sulla metropoli. La stanza trema, il cielo s'incendia, un meteorite si schianta sulla città, le luci sfarfallano, l'onda d'urto avanza, polverizza i palazzi sulla sua strada, si alzano urla, c'è chi cade a terra, il datore scappa... Ma quella non è una finestra, è solo un televisore ad alta definizione.

Ora immagino che, mentre il televisore trasmette quel falso video, un meteorite cada realmente sulla città spargendo distruzione. Capisco da come reagiscono che le vittime dello scherzo credono davvero che un meteorite stia per distruggere la città. E in effetti hanno un buon motivo per crederlo: stanno vedendo quelle cose coi loro occhi. Per di più, è vero: stanno vedendo un falso filmato ma, mentre lo vedono, un vero meteorite si sta realmente schiantando sulla città.

Quindi, il contesto descritto soddisfa tutti e tre i requisiti dell'analisi di conoscenza: le vittime credono giustificatamente una cosa vera... epperò, non la chiamerei conoscenza: la loro credenza è causata da un inganno, che questa sia vera è solo fortuna (o, meglio, sfortuna).

Si possono costruire esempi a piacere. Penso al regista Orson Welles che leggeva alla radio dell'atterraggio di una nave aliena (Welles 1938): alcuni ascoltatori (ma meno di quanto si dice) lo credevano, tanto che si erano messi in macchina pronti alla fuga, e avevano anche dei motivi per crederlo visto che stavano ascoltando una fonte che si presuppone affidabile; e ora immagino che, per caso fortuito, degli alieni stessero realmente visitando la Terra in quell'istante... Anche qui: ho una credenza vera giustificata, ma non la direi conoscenza.

Questi sono i *casi di Gettier*. Gettier difficilmente viene citato tra i grandi filosofi viventi, ma nessuno più di lui ha influenzato il discorso sul concetto di conoscenza (Perry 2015). Il suo incredibile merito è quello di aver pubblicato una ricerca di due pagine (Gettier 1963) in cui espone esempi di questi casi per nulla originali e noti sin dal medioevo (e.g. Pietro Mantovano 1492). Ma se per i filosofi medievali non erano problematici, per i filosofi analitici sono un enorme stimolo, perché mostrano che l'attuale analisi di conoscenza è, nel migliore di casi, incompleta, e nel peggiore errata.

Si può obiettare che il problema è mio, e non dell'analisi? Cioè, posso dire che l'analisi è corretta, e quindi che quella dei casi di Gettier è vera conoscenza, e sono io che sbaglio nel non riconoscerla come tale? Si può dire, ma sarebbe una risposta anti-analitica.

L'analisi di un termine è *descrizione*, non *normazione*, rispondere così vorrebbe dire imporre un nuovo concetto di conoscenza diverso da quello del senso comune, creare arbitrariamente una nuova parola e non scoprire il senso di quella normalmente usata. Se si vuole creare una teoria normativa che senso ha tentare un'analisi in primo luogo? Wittgenstein piange vedendo certe cose. Come un fisico parte dalle osservazioni e non può dire che i risultati dei test sono sempre errati, così il filosofo analitico parte da intuizioni sull'uso delle parole e non può dire che l'intuizione linguistica è sistematicamente errata.

La conoscenza deve essere indefettata 2: perché affidabilità e infallibilità sono di troppo

Nelle pagine seguenti mi rifaccio ad un diverso esempio preso da Russell 2009: guardo l'orologio, vedo che sono le due, e sono davvero le due, però l'orologio è fermo. Quindi: credo che sono le due, è vero, e sono giustificato a crederlo perché guardare l'orologio è un normale giustificatore della credenza sull'ora, ma non pare conoscenza perché, essendo l'orologio fermo, lo credo *per i motivi sbagliati*.

Cosa c'è che non va? Non paiono esservi problemi con le nozioni di verità e di credenza. Allora: o non sono giustificato come credo, e ho bisogno di una definizione più stretta di giustificazione; oppure sono davvero giustificato, ma devo aggiungere una quarta clausola; oppure sono davvero giustificato, ma la giustificazione non è una componente della conoscenza e devo sostituirla con una nuova clausola.

È evidente qual è la soluzione: devo trovare il modo di specificare che conosco l'orario se: lo credo, è corretto, e, se lo credo guardando l'orologio, l'orologio non deve essere fermo. Per capire come arrivarci mi chiedo: *perché* non posso avere conoscenza con un orologio fermo?

Posso escludere subito alcune versioni basiche dell'*affidabilismo*:

AF: ho conoscenza solo se la credenza è formata tramite un *processo conoscitivo affidabile*

(dico «solo se» e non «sse» perché l'affidabilità, in queste teorie, è una condizione necessaria ma non sufficiente) (Goldman 1986, Bach 1985). Infatti, guardare l'orologio è *generalmente* un processo affidabile per conoscere l'ora, il fatto che ogni tanto possa essere fermo non lo rende inaffidabile; stessa cosa per il meteorite, la vista è un processo molto affidabile, quindi non può essere questo il motivo per il quale non è conoscenza.

Un'altra naturale diagnosi è che sto cercando di derivare la conoscenza da una premessa sbagliata, cioè è come se stessi pensando:

P1 vedo che l'orologio segna le due;

P2 non c'è motivo per il quale l'orologio possa sbagliarsi;

C quindi sono le due;

ma P2 è falsa (in questo caso l'orologio non sbaglia, ma avrebbe potuto sbagliare). È dunque sufficiente aggiungere una clausola *no false premise*

NFP: ho conoscenza solo se la credenza non discende da premesse false.

Un grande problema di questa teoria sta nel suo presupporre che la conoscenza derivi sempre e solo da ragionamenti; in questo caso, però, o non è così, o il ragionamento è mal rappresentato: dal fatto che l'orologio segna le due io deduco direttamente che sono le due, senza passare per una P2 di questo tipo.

Allora si può obiettare che è proprio P1 ad essere errata perché è da intendere come «vedo un orologio *funzionante*», non lo è quindi sono ingiustificato. Ma si tratta di un'interpretazione tendenziosa: che l'orologio sia funzionante, o non funzionante, o addirittura sia un tatuaggio, la mia percezione è sempre uguale, in tutti e tre i casi l'interpretazione piana «ho l'esperienza visiva di un orologio» rimane vera.

Richiedere la verità dell'interpretazione «vedo un orologio funzionante», cioè affermare che la percezione può giustificarmi solo se *veridica* (solo se vedo proprio ciò che credo), o richiedere la verità della P2 «non c'è motivo per il quale l'orologio possa sbagliarsi», porta ad un secondo grande problema della teoria: quello dell'*infallibilità*.

L'idea di una intrinseca infallibilità della conoscenza può essere intuitiva. Anche un orologio rotto segna l'ora giusta due volta al giorno: e allora perché guardandolo non posso conoscere l'ora due volta al giorno? Perché viene a mancare il funzionamento che giustifica l'accordo sistematico tra l'ora reale e la posizione delle lancette, diventano due cose disconnesse. Cade così il motivo per il quale guardare l'orologio è un metodo affidabile di conoscenza: guardarlo non può più giustificare la mia credenza perché, se non fosse stata davvero l'ora segnata, l'avrei creduta comunque. Non basta l'affidabilità del metodo, serve che non possa giustificare credenze false, deve essere *immune all'errore*.

Ci sono anche motivi per ritenere che, in effetti, i casi di Gettier possano essere risolti *solo* richiedendo giustificatori infallibili. Ecco la ricetta per creare un caso di Gettier (Zagzebski 1994):

1. individuo una credenza (giustificata ma) falsa;
2. modifico il contesto in modo tale che la credenza risulti vera per pura fortuna;
3. informare a 180 °C per trenta minuti, e il caso di Gettier è pronto.

Se è così, allora l'unico modo per evitare ogni caso di questo tipo è porre una definizione di giustificazione così forte da non permettere mai credenze (giustificate) false.

Posso provare ad esporre questo *principio di infallibilità* in più modi:

INF 1: ho conoscenza solo se la mia credenza (o il suo giustificatore) è dipendente dalla sua verità;

INF 2: ho conoscenza solo se, se la credenza non fosse vera, non la crederei (o non sarei giustificato a crederla);

INF 3: ho conoscenza solo se si dà un contesto tale che non avrei creduto quella credenza (o non sarei stato giustificato a crederla) se questa non fosse stata vera.

Nessun filosofo a me noto ha esposto un principio in questi termini, ma ne esistono vari di simili, per esempio quello alla base della teoria *truth-tracking* di Robert Nozick, uno dei più interessanti e spirituali filosofi contemporanei (Nozick 1981, Dretske 1971, cfr. Sosa 1999).

Il problema di questo principio è che è troppo restrittivo. Una conoscenza è tale se è vera, ma chiedere un giustificatore così forte da non poter mai giustificare credenze false (o comunque, se si rifiuta il criterio della giustificazione, un contesto tale da farmi credere un certo fatto solo se vero) va contro il senso comune: se l'analisi di conoscenza proposta è troppo liberale perché riconosce come conoscenza anche casi che non sono tali, aggiungere questo principio la rende invece troppo restrittiva, definisce conoscenza troppo poco rispetto a quello che si considera normalmente così. Le migliori teorie scientifiche prodotte dall'umanità sono fallibili, il metodo scientifico potrebbe portare a credenze false, ma nessuno direbbe per questo che, se invece sono vere, non sono conoscenza.

Una variante più permissiva è la *teoria causale* di Goldman (1967), uno dei maggiori epistemologi viventi, cioè:

TC: ho conoscenza solo se c'è una connessione causale tra la credenza e la sua verità.

Questo permette di rendere merito della conoscenza scientifica ed è compatibile con tutti i casi in cui il giustificatore è una percezione, epperò pare un principio ancora troppo rigido. Io posso avere conoscenze sul futuro, io so che a settembre riapriranno le scuole, ma una verità futura non può causare una credenza presente. Io credo che la medicina mi farà stare bene, e lo credo perché *tutte le altre volte* è stato così: è statistica, non implica che *ogni volta* mi farà stare bene né si collega alla verità del fatto che *ora* mi farà stare bene, eppure è conoscenza. Il mio credere agli elettroni è giustificato da una certa osservazione causata dagli elettroni, ma il mio credere è *causato* esclusivamente dalla speranza nel successo della scienza (lo credevo già prima di essere giustificato, avessi sperato altro avrei comunque creduto diversamente): in questo caso la credenza non è causata dalla verità, eppure, se si accetta la componente della giustificazione, in virtù di questa è comunque conoscenza.

La conoscenza deve essere indefettata 3: un semi-esternalismo

L'idea dell'infallibilità è giusta: affinché una credenza vera sia conoscenza è necessario che sia giustificata *in un certo modo*; non è sufficiente che la percezione mi giustifichi una credenza vera, è necessario che la giustifichi tramite la visione dell'oggetto della mia credenza (e non di altro). Ora la strategia è quella di indebolire il principio per ottenerne un altro più inclusivo ma che continui a rifiutare i casi di verità per pura fortuna.

Posso provare, seguendo Lehrer, a passare dall'infallibilità all'*indefettibilità* (Lehrer & Paxson 1969). È defettibile ciò che *può venir meno*, quindi una giustificazione è indefettibile se non può essere invalidata. Questo io non lo intendo in senso assoluto, intendo che non può essere invalidata dalla conoscenza di alcun altro elemento effettivamente presente nel contesto (dunque, non indefettibile ma indefettata). Quindi, ecco il *principio di indefettità*:

PI: ho conoscenza solo se non ci sono elementi (noti o meno) che invalidano il mio giustificatore (cioè che mettono in dubbio la sua affidabilità).

Per esempio, nel caso dell'orologio, anche se il guardarlo è generalmente un mezzo affidabile, il fatto che sia fermo lo rende non più tale, e questo lo noto perché, se io sapessi che l'orologio è fermo, non mi riterrei più giustificato.

Sia chiaro che questo non è da intendere come un criterio ulteriore alla giustificazione, ma come una sua componente: sarebbe infatti ben strano dire «sono giustificato anche se esiste un elemento che annulla la mia giustificazione (e per questo non ho conoscenza)», se questo elemento esiste non sono giustificato a prescindere.

Se si accetta ciò, PI implica una forma di esternalismo della giustificazione (cfr. Bonjour 1980). Come nel caso della Terra* gemella il significato non dipende dagli stati mentali ma dal rapporto con l'ambiente, così il mio essere giustificato o meno non dipende da ciò che posso pensare o sapere ma dagli elementi del contesto. Se io guardo l'orologio e vedo che segna correttamente le due, ci sono due possibilità: o è fermo e segna l'ora giusta per coincidenza, o funziona e segna l'ora giusta in virtù di ciò. Nel primo caso non c'è conoscenza, nel secondo sì, e questo non dipende dal mio sapere se l'orologio funziona, ma dal suo effettivo funzionamento.

Questo non significa però che la giustificazione sia solo esterna: può ben avere sia elementi esterni che interni. Da un lato, se io continuo a guardare l'orologio e mi accorgo che non va avanti, posso dire «ah, credevo di essere giustificato, e invece non lo ero»: questo è un elemento esterno perché la mia giustificazione non viene invalidata nel momento in cui vengo a sapere internamente che l'orologio è fermo, bensì non sono mai stato giustificato. Dall'altro lato, se uno mi chiede quante persone ci sono in un palazzo e io rispondo «1255 perché sono le 12:55», la mia credenza può essere giustificata solo se ho accesso interno ad una teoria che collega l'orario al numero di persone, non devo verificare ogni volta questo collegamento ma non è neanche sufficiente che vi sia senza essermi noto: questo deve essere un elemento interno, non posso dire «non sapevo di essere giustificato».

Così, ad un primo livello, io credo correttamente che sono le due, e la mia giustificazione è il guardare l'orologio. Perché guardare l'orologio è un giustificatore? Questo porta alla *metagiustificazione*: perché l'orologio si muove in modo tale da indicare l'ora corretta (non è necessario, come nel caso di una meridiana, un rapporto *causale* tra l'ora indicata e l'ora reale, ma deve esservi una connessione). Questa è comunque una conoscenza, quindi ha un aspetto interno

(devo credere che l'orologio si muova così, e devo avere motivi per crederlo) e un aspetto esterno (deve essere vero). Se l'orologio non si muove il metagiustificatore risulta falso, e questo è sufficiente, anche a mia insaputa, a invalidare la giustificazione; cioè un elemento mette in dubbio l'affidabilità di un giustificatore quando nega la sua metagiustificazione. Quindi gli elementi esterni non possono giustificarmi, ma la verità dei metagiustificatori è necessaria alla mia piena giustificazione. Questo ricorda il principio *no false premisses*, ma ci sono delle differenze: il metagiustificatore mi giustifica nell'usare una credenza come giustificatore, ma non è il giustificatore di questa credenza, non appare necessariamente all'interno di ragionamenti, né mi permette sempre di inferire la credenza giustificata.

Considero ora questo esempio da Lehrer & Paxson 1969:

- vedo Tom (o un uomo uguale a lui) rubare un libro;
- a mia insaputa, sua madre testimonia ad un processo che Tom ha un fratello gemello cleptomane;
- sempre a mia insaputa, la madre di Tom è una bugiarda patologica, perché in realtà non ha altri figli.

Questo contesto è problematico se si ritiene che una credenza sia indefettibile: in tal caso, il vedere Tom sarebbe un giustificatore defettibile perché, se sapessi ciò che sua madre ha testimoniato, non mi riterrei giustificato, ma in realtà la mia è comunque conoscenza perché sua madre è una bugiarda, quindi l'indefettibilità non è una condizione necessaria. Io però, parlando invece di un giustificatore indefettato, devo considerare il complesso di tutti gli elementi effettivamente esistenti. Perché conoscere la testimonianza mi farebbe credere non-giustificato? Non per ciò che dice la madre, ma perché crederci nell'esistenza di un suo fratello gemello. Io sono giustificato se nel contesto non ci sono elementi che mettono in dubbio l'affidabilità dei miei giustificatori: quindi, se il fratello esistesse, per essere giustificato dovrei prima accertarmi dell'essere in grado di distinguerli. Ma per fortuna il fratello non esiste, quindi nessun elemento esistente mette in dubbio l'affidabilità delle mie percezioni, quindi rimango giustificato.

Saul Kripke (2011, ma basato su Goldman 1976) offre un altro esempio contro la teoria di Nozick 1981 che è problematico anche per l'indefettibilità. C'è una regina con una gran passione per la campagna e i suoi fienili, quando viaggia vuole sempre vederne il più possibile, così i sudditi, per compiacerla, hanno riempito una strada di fienili falsi (hanno solo la facciata); i fienili falsi sono tutti verdi, ma esiste un unico fienile vero ed è rosso. Un pellegrino capita davanti a quell'unico fienile e pensa «qui c'è un fienile rosso». È una credenza vera e giustificata in modo indefettato: infatti, avesse una conoscenza completa del contesto, riconoscerebbe che non c'è alcun fienile falso rosso che può invalidare la giustificazione. Ciò nonostante, se io so che lì «c'è un fienile rosso», saprò anche che «c'è un fienile»; ma questa seconda credenza non è indefettata, l'esistenza di fienili falsi dovrebbe invalidare la giustificazione, eppure io so, quindi l'indefettibilità non è una componente della conoscenza.

Ci sono varie risposte a questa argomentazione. Una è questa: è possibile che il fatto che «qui c'è un fienile» sia implicata da una credenza giustificata in modo indefettato sia sufficiente a giustificarla anche se non-indefettata di suo. Ho anche un'altra, più convincente risposta, ma prima devo parlare un po' dello scetticismo...

La conoscenza deve essere aperta: i contesti scettici

La Vida es Sueño, di Calderón de la Barca (1635). Basilio, re di Polonia appassionato di astrologia, vaticina che suo figlio Sigismondo diventerà un tiranno sanguinario, lo rinchiude così in una torre. Pentito, il re decide di dare una possibilità al figlio per vedere se può spezzare le ali del destino, e così, mentre dorme, lo fa portare alla reggia. Sigismondo, disabituato alla civiltà, finisce per comportarsi davvero in modo tirannico, così il re lo fa riaddormentare e riportare alla torre, dove si risveglierà convinto che sia stato tutto un sogno.

Quando sogno, non so di sognare. Vedo cose che credo essere reali ma non lo sono, tutto ciò che credo in sogno non può costituire conoscenza, sono tutte credenze o false, o vere per sola fortuna e quindi ingiustificate. Ma io tutt'ora non posso escludere di star sognando, non posso escludere che tutto ciò che mi circonda sia un'immaginazione: come posso affermare di avere conoscenza?

Il sogno e situazioni simili (come il demone di Cartesio 1641) sono *contesti scettici*. Il meccanismo di base è simile a quello dei casi di Gettier: dalla possibilità dell'errore si deduce l'impossibilità della conoscenza. Ma mentre i casi di Gettier descrivono situazioni specifiche con elementi che invalidano la giustificazione solo di alcune credenze, i contesti scettici immaginano mondi con elementi che invalidano ogni possibile giustificazione e rendono così la conoscenza sempre e comunque impossibile.

Un contesto esemplare di questo tipo è quello dei cervelli in vasca di Putnam (cfr. Sosa 1999). Anzitutto, per definizione, dato che in un contesto scettico non posso sapere alcunché (e.g. in sogno non posso acquisire alcuna conoscenza), anche se tale contesto non si desse, non potrei sapere che non si dà (anche se non sogno, non posso sapere che non sto sognando). Quindi io non posso sapere di non essere un cervello in vasca, e usando il linguaggio visto per il paradosso di Fitch (con la *s* ad indicare il contesto scettico) scrivo:

$$\mathbf{P1} \quad \neg K \neg s$$

Sempre per definizione, se io sono un cervello in vasca, non ho un corpo (al di là del mio cervello). Quindi, se ho un corpo (fatto che indico con *p*), non sono un cervello in vasca. Dato che questa verità è analitica, sono automaticamente giustificato a crederla, quindi posso dire di conoscerla:

$$\mathbf{P2} \quad K(p \rightarrow \neg s)$$

Conoscendo ciò, è intuitivo che, se so anche di avere un corpo, crederò di non essere un cervello in vasca; quindi, indicando con *B* la credenza:

$$\mathbf{P3} \quad (Kp \wedge K(p \rightarrow \neg s)) \rightarrow B \neg s$$

C'è ora un principio accettato da quasi tutte le epistemologie e che fa da base alla filosofia di Cartesio, il *principio di chiusura epistemica*. Cosa fa Cartesio? Partendo da un contesto scettico, individua un'unica credenza indubitabile (il *cogito*), e quindi una conoscenza, e accetta come ulteriore conoscenza tutto ciò che riesce a implicare da questa. Cioè Cartesio mi dice che io posso ampliare la mia conoscenza semplicemente indagando le implicazioni delle conoscenze acquisite, cioè che tutte le credenze che io so essere implicare da una conoscenza sono a loro volta conoscenza:

$$\mathbf{PC} \quad (Kp \wedge K(p \rightarrow q) \wedge Bq) \rightarrow Kq$$

che, adattato al contesto dei cervelli in vasca, diventa:

$$\mathbf{P4} \quad (Kp \wedge K(p \rightarrow \neg s) \wedge B\neg s) \rightarrow K\neg s$$

Suppongo ora di sapere che io ho un corpo (o qualsiasi altro fatto banale che neghi il mio essere un cervello in vasca, come l'averne qualche conoscenza). Posso congiungere tale conoscenza con P1 per dire che so di avere un corpo e non so di non essere un cervello in vasca:

$$\mathbf{P5} \quad Kp \wedge \neg K\neg s$$

Ma, dandosi P2 e P3, in virtù di P4, P5 diventa

$$\mathbf{P6} \quad K\neg s \wedge \neg K\neg s$$

cioè io so e non so di avere un corpo, il che è assurdo. Quindi la supposizione è falsa, e devo negarla:

$$\mathbf{C} \quad \neg Kp$$

Questa è la conclusione del ragionamento, che però è chiaramente inaccettabile: io so di avere un corpo, so di sapere qualcosa, e quindi devo riuscire a rendere merito del mio sapere di non essere un cervello in vasca. Dunque devo rifiutare una delle premesse, ma quale?

Anche se ci sono più possibilità d'azione, in virtù del PI che ho proposto, devo rifiutare PC (e quindi P4). Cosa succede rifiutando PC? Che il mio sapere che ho un corpo non implica il mio sapere che non si dà il contesto scettico, e quindi non c'è più contraddizione col fatto che io non posso sapere che un contesto scettico non si dà.

Altro esempio di come PI senza PC riesce a gestire un gran numero di contesti scettici, il contesto del sogno (come in Aristotele 2010 1011a-6):

P1 io non posso sapere che non sto sognando;

P2 se sto sognando, nulla di ciò che credo è conoscenza;

P3 per avere conoscenza, devo dimostrare che non sto sognando;

C ma per P1 non posso dimostrarlo, quindi non posso conoscere alcunché.

Cosa dovrei dire, accettando l'indefettità? Internamente, io sono giustificato a credere che questa con cui sto scrivendo sia la mia mano. Esternamente, se davvero sto sognando, davvero tutte le mie giustificazioni sono invalide, anche questa. Ma se non sto sognando, allora non c'è nulla che la invalidi. Affinché la mia giustificazione sia valida non è necessario che io sia in grado di escludere il contesto scettico come vorrebbe P3, è sufficiente che in effetti non si dia. E questa è l'apertura epistemica: il fatto che io sappia qualcosa implica che il contesto scettico non si dà, ma non implica che io lo sappia.

Ma perché non posso accettare PI e PC insieme? Riprendo ora l'esempio dei fienili di Kripke. Io so che

F1 Qui c'è un fienile rosso

e so anche che da questa si implica che

F2 Qui c'è un fienile

Io so F1, so che $F1 \rightarrow F2$, credo F2, quindi, dato PC, io conosco F2. F1 è indefettata perché non esistono elementi che invalidino la sua giustificazione; questo non significa, però, che non esistano elementi del genere anche per tutto ciò che posso derivare da F1: e in effetti, F2 non è indefettata, l'esistenza dei fienili falsi blocca la sua giustificazione. Quindi per PC io conosco F2, ma per PI non lo conosco perché non sono giustificato; quindi PC e PI sono incompatibili, se accetto uno devo rifiutare l'altro.

Credo che questa conclusione sia corretta, e poiché io accetto PI, devo rifiutare PC. Non per questo però devo ingoiare il rospo (o, come dicono i filosofi inglesi, mordere il proiettile) e riconoscere che, pur conoscendo F1, non conosco F2. Il punto è questo: qual è il giustificatore di F2? Non il mio vedere un fienile (com'è invece per F1), ma il derivare logicamente l'esistenza del fienile dall'esistenza di un fienile rosso. Questo mio giustificatore logico non viene affatto invalidato dall'esistenza di fienili falsi, quindi F2 è effettivamente indefettata.

Si può obiettare che, se è così, allora PC risulta compatibile con PI, perché tutte le credenze che derivano logicamente da una credenza indefettata vera risultano essere conoscenze. Ma non è così, perché anche l'implicazione logica potrebbe essere invalidata dall'esistenza di certi elementi del contesto: come detto, certi effetti della fisica quantistica potrebbero portare a rivedere la logica; oppure potrei aver assunto sostanze stupefacenti e in tal caso, come devo prima essere sicuro di poter distinguere Tom da suo fratello gemello per essere giustificato dalla mia vista (a prescindere dalla loro effettiva distinguibilità), così devo assicurarmi di riuscire a ragionare normalmente prima di poter essere giustificato dal notare un'implicazione logica (a prescindere dalla sua effettiva correttezza formale). Certo, forse si può creare un nuovo tipo di principio di chiusura adatto all'indefettità, ma questo tipo standard non è accettabile.

Un'altra obiezione è che PC è intuitivo e indiscutibile, e non potendo essere rifiutato, è PI a dover essere escluso. Ma non credo che la questione sia così semplice.

Dretske, per esempio, propone numerosi esperimenti mentali che mettono in dubbio la validità di PC. Se vedo una zebra, sono giustificato a credere che lì vi sia una zebra, e credendo ciò implicherò che lì non c'è invece un zebra-robot molto realistica; epperò, non avendo vere prove, non posso essere giustificato a credere ciò (Dretske 1970). Oppure, se io scelgo di acquisire conoscenza tramite percezione, e percepisco un certo fatto, e so che questo fatto ne implica un altro, comunque non percepisco anche questo altro fatto, quindi non lo conosco per percezione (e potendo ripetere l'esempio per ogni altro mezzo di conoscenza, la conclusione risulta di validità generale). Ancora: io so di aver fatto colazione stamattina, e da questa conoscenza posso implicare che il passato esiste (cioè esistono fatti passati che rendono vera tale frase); eppure, il solo sapere di aver fatto colazione non può certo portarmi ad avere conoscenza di una così profonda verità metafisica, servono ben altri giustificatori (Dretske 2003, 2005).

Ecco un altro esempio più complesso, basato su Comesaña (2014).

Dal fatto che io vedo correttamente un tavolo (p) sono giustificato a dedurre che qui c'è un tavolo (q):

P1 $p \rightarrow q$

Suppongo che P1 sia vera, e quindi che vi sia davvero un tavolo. Indicando la disgiunzione «o» col simbolo \vee , esiste una tautologia classica (una proposizione sempre vera) per la quale

$$\mathbf{T1} \quad ((p \rightarrow q) \wedge q) \rightarrow (q \rightarrow (q \vee \neg p))$$

cioè, se si dà P1, e c'è un tavolo, da quest'ultimo fatto posso dedurre che, o c'è un tavolo, o non sto davvero vedendo un tavolo:

$$\mathbf{P2} \quad q \vee \neg p$$

Poiché P2 è del tutto equivalente a P1 (è solo un altro modo di scriverlo), anche da P1 potrei dedurre P2. Dunque, suppongo di conoscere P1, di sapere che P1 implica P2, e di credere P2. Accettando PC posso scrivere:

$$\mathbf{P3} \quad (KP1 \wedge K(P1 \rightarrow P2) \wedge BP2) \rightarrow KP2$$

Poiché io conosco P2, devo essere giustificato a crederlo. Ma: cos'è che mi giustifica a credere P2? Dato che P2 è stato dedotto da q, è q a giustificarla; ma q è a sua volta giustificata da p, quindi q non ha potere giustificante di per sé ma bensì è p a giustificare P2; quindi, indicando la giustificazione con J, scrivo:

$$\mathbf{P4} \quad p \rightarrow J(q \vee \neg p)$$

La mia supposizione ora è che io sia giustificato a credere P2 (in virtù di p, non essendovi altri giustificatori papabili). Intuitivamente, se sono giustificato a credere P2, sono anche giustificato a *non* credere il suo contrario, cioè sono giustificato a non credere che io veda un tavolo senza che questo vi sia (non fosse così, potrei credere sempre di essere allucinato):

$$\mathbf{P5} \quad J(q \vee \neg p) \rightarrow J \neg (p \wedge \neg q)$$

Ovviamente se una congiunzione «a e b» è vera, si dà a e si dà b, e sono quindi singolarmente deducibili dalla congiunzione. Questa è un'altra tautologia classica:

$$\mathbf{T2} \quad (p \wedge \neg q) \rightarrow p$$

Sulla base di P5 e T2, da P4 ottengo:

$$\mathbf{P6} \quad p \rightarrow J \neg p$$

Ma che un giustificatore giustifichi a *non* credere al giustificatore medesimo è assurdo, quindi devo negare la supposizione:

$$\mathbf{C} \quad \neg J(q \wedge \neg p)$$

L'idea è che, dato che P4 discende da P3, se devo negare P4, devo negare anche P3, cioè PC. Ovviamente, questo non è un obbligo: P4 non discende *solo* da PC, ma anche dalle altre premesse (e da una serie di regole di deduzione implicite), quindi posso mantenere PC e rifiutare altri punti. Rifiutare PC, però, rimane una soluzione.

Riconosco che il rifiuto della chiusura epistemica deve essere molto meglio difeso, devo fare molte più ricerche, molto più specifiche, su questo singolo argomento. E credo anche che controesempi al principio di indefettività siano possibili. L'analisi della conoscenza potrebbe essere molto complessa, forse prevede delle disgiunzioni (potrebbe richiedere una giustificazione molto forte o una giustificazione debole più una quarta clausola). Ma credo pure, in linea di massima, che esternalismo e fallibilità della conoscenza siano caratteristiche da mantenere.

Se ogni mia impresa intellettuale è fallibile, in ogni conoscenza c'è un elemento di speranza: poiché io non posso sapere se un certo contesto scettico si dà o non si dà, e se si dà non posso sapere alcunché, posso solo agire con la speranza che non si dia. La speranza è il fondamento della conoscenza.

Le due tesi 1: fondazionalismo e coerentismo

Il Barone di Münchhausen, grande epistemologo, rimasto intrappolato nelle sabbie mobili del dubbio, è riuscito a salvarsi tirandosi su per i capelli.

Come faccio a sapere che ho le mani? Perché le vedo con gli occhi. E come faccio a sapere che gli occhi vedono il vero? Perché, quando guardo una mano, vedo una mano e non altro. Ma questo è circolare, è come dire che la Bibbia è infallibile perché la Bibbia dice di esserlo, non può giustificarmi. Allora: perché conosco la fisiologia e l'ottica. E come faccio a conoscerle? Ho letto libri. E come faccio a sapere che dicono il vero? Posso andare avanti a farmi domande all'infinito, nessun può calcolare all'indietro infinite giustificazioni, ad un certo punto devo pur fermarmi... ma se mi fermo, arriverò ad una affermazione creduta perché sì, quindi comunque non può giustificarmi.

Quello che segue è il *trilemma di Münchhausen*, un ragionamento che Hans Albert (1968) fa risalire a Popper (2010), il quale lo fa risalire a Fries (1807), ma che risale in realtà a Sesto Empirico, che dice risalire invece ad Agrippa (Sesto Empirico 1999), il quale si stava sicuramente rifacendo ad Aristotele (2003, e.g. 72b ss.):

- ogni credenza è da giustificare, ogni giustificatore è esso stesso una credenza, un giustificatore può giustificare solo se giustificato a sua volta;
- 1. quindi dovrebbero esistere infiniti livelli di giustificazione, ma nessuno può calcolare giustificazioni di giustificazioni all'infinito, quindi nessuno può essere giustificato;
- 2. allora ci si dovrebbe fermare ad un giustificatore senza giustificazione, una *credenza basilica* giustificata e giustificante di per sé, ma questo è dogmatismo, nessuna credenza ingiustificata può giustificare;
- 3. allora si dovrebbe trovare un insieme di giustificatori che si giustificano l'un l'altro in circolo, ma neanche il ragionamento circolare può giustificare.

Sembra proprio che la giustificazione sia impossibile... Questo trilemma è la base dello scetticismo, se non si vuole essere scettici bisogna negare una delle affermazioni precedenti:

1. chi nega che sia impossibile regredire all'infinito a giustificazioni sempre più profonde sostiene l'*infinitismo*;
2. chi nega che sia impossibile individuare credenze basiliche giustificanti senza essere giustificate da un'altra credenza sostiene il *fondazionalismo*;

3. chi nega che il ragionamento circolare non possa giustificare sostiene il *coerentismo*.

Delle tre teorie, l'infinitismo è quella con meno fan. Gli epistemologi sono motivati dal rifiuto dello scetticismo, ma spesso lo sono anche dal rifiuto della necessità di un regresso infinito. Ciò nonostante, nell'ultimo periodo stanno iniziando a prendere la tesi più sul serio.

L'infinitismo accetta che una giustificazione non può essere né arbitraria né circolare, quindi, a meno di negare la necessità della giustificazione, bisogna accettare il regresso infinito. Questo non vuol dire che bisogna davvero calcolare infinite giustificazioni (anche gli infinitisti riconoscono che è impossibile), è sufficiente che *sarebbe possibile* farlo, è sufficiente che esista effettivamente un set di infinite ragioni per credere ciò che si crede (Klein 1998).

Io penso che l'ideale per l'infinitismo sarebbe quello di sviluppare una teoria che spieghi perché, se una credenza ha certe caratteristiche, sarei in grado di portare giustificazioni all'infinito: io so che i numeri sono infiniti anche se non ho mai provato a calcolarli tutti perché la regola «aggiungi uno», applicabile ad un numero qualsiasi, può poi essere riapplicata al suo risultato; similmente, dovrei trovare un modo di generare giustificazioni in modo *ricorsivo*. Se non riesco a dimostrare qualcosa del genere, allora la giustificazione rimane totalmente esterna, non posso sapere se sarei in grado oppure no di portare giustificazioni su richiesta all'infinito.

Per questo ho scelto di concentrarmi solo su fondazionalismo e coerentismo.

Analizzando i problemi di Gettier ho già individuato una componente della giustificazione, l'indefettibilità, la quale però non è sufficiente: è vero che non devono esservi elementi esterni tali da annullare la mia giustificazione, ma devo anche sapere quali sono le ragioni interne che mi giustificano a credere una certa cosa, e questo è ciò che fondazionalismo e coerentismo aggiungono alla teoria.

Nel fondazionalismo una credenza o è giustificata da un'altra credenza, o è basilica; quindi alla fine ogni catena di giustificazioni si ferma ad uno strato di credenze basiliche. Lungi dall'essere un dogma ingiustificato, la credenza basilica è *giustificata in sé, auto-giustificante*, o giustificata *non-inferenzialmente*, non si risale oltre perché la sua giustificazione non sta in altre credenze: per esempio, posso sostenere che il mio credere che qui vi sia un tavolo è giustificato direttamente dalla mia esperienza sensoriale del tavolo (non dal mio credere di vederlo).

Il coerentismo si contraddistingue per la *tolleranza* dei processi circolari in cui due o più credenze si giustificano una con l'altra. Il fondazionalismo non riconosce a processi di questo tipo un potere di giustificazione, il coerentismo sì: ciò non vuol dire che un sistema coerentista *debba* prevedere processi circolari, possono anche non esserci, ma se ci sono ciò non distrugge la giustificazione.

Come la controparte in teoria della verità, il coerentismo afferma che una credenza non può essere giustificata di per sé, bensì è giustificata solo nella misura in cui appartiene ad un sistema di credenze con determinate caratteristiche: è l'insieme, non la credenza presa in isolamento, l'oggetto della giustificazione. Una credenza può essere detta giustificata se è coerente con un certo sistema di credenze, ed è tanto più giustificata tanto più è coerente il sistema a cui appartiene. Questo implica, in primo luogo, che il coerentismo ragiona in continuo, ci sono credenze più o meno giustificate, e il loro grado di giustificazione è relativo ai vari sistemi in cui le si pone. In secondo luogo, implica che solo la relazione tra credenze può giustificare, non lo possono fare altre cose come le percezioni.

Le due tesi 2: una necessaria introduzione

Mi aiuta rappresentare visivamente la forma delle due teorie:

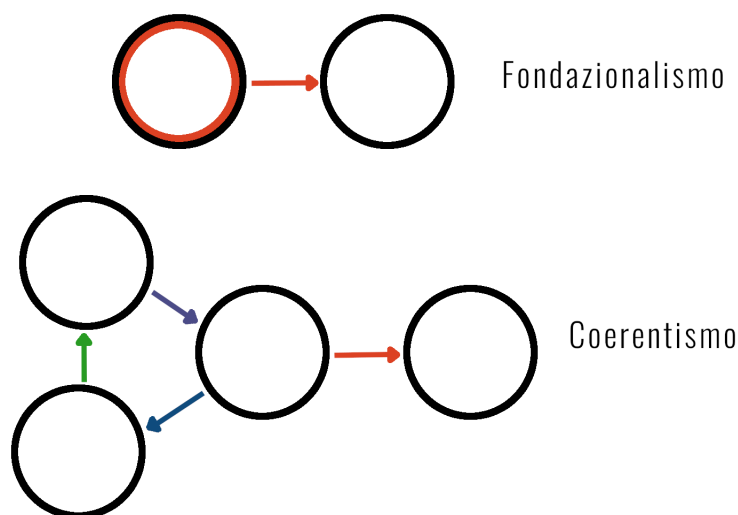


Figura 2

I cerchi rappresentano una o più credenze (il cerchio colorato la credenza basilica), le frecce il rapporto di giustificazione. Quest'ultimo è spesso un rapporto d'*inferenza*: una credenza ne giustifica un'altra nel senso che dalla prima si può inferire la seconda, cioè esiste un ragionamento che, premessa la prima, porta a concludere la seconda. Individuo così i tre classici tipi d'inferenza: *deduzione, induzione, abduzione* (cfr. Aristotele 2003 e Peirce 2011):

Fatto: Socrate è un uomo.

Regola: se uno è un uomo, allora è mortale.

Conclusione: Socrate è mortale.

Questa è una deduzione, il tipo di ragionamento tipico della matematica e della filosofia: da due premesse vere si ricava una conclusione certa.

Fatto 1: Socrate è un uomo.

Fatto 2: Socrate è mortale.

Conclusione: se uno è un uomo, allora è mortale.

Questa è una induzione, il tipo di ragionamento della scienza. Richiede osservazioni ripetute: se vedo un milione di uomini, e vedo che sono tutti mortali, la generalizzazione «ogni uomo è mortale» sarà molto probabile, anche se non può mai essere certa.

Fatto: Socrate è mortale.

Regola: se uno è un uomo, allora è mortale.

Conclusione: Socrate è un uomo.

Questa è l'abduzione, il tipo di ragionamento utilizzato per fare ipotesi sulle spiegazioni dei fatti osservati. È ovvio che ogni fatto può avere infinite spiegazioni, ogni uomo è mortale ma anche ogni cane gatto coniglio, quindi la conclusione è sempre dubbia; epperò, è anche la più creativa, è il tipo di ragionamento che più può aumentare la mia conoscenza di Socrate.

Se devo rappresentare un ragionamento di questo tipo, nello schema, tre cerchi non possono rappresentare sempre e solo isolatamente le tre credenze «Socrate è un uomo», «Socrate è mortale», «ogni uomo è mortale»: io inferisco le conclusioni partendo da due premesse, da una sola premessa non posso inferire alcunché. I cerchi devono rappresentare anche unità concettuali più complesse; per esempio, il sistema di credenze «Socrate è un uomo, è mortale, e ogni uomo è mortale» lo rappresenterei meglio con tre elissi colorate, così:

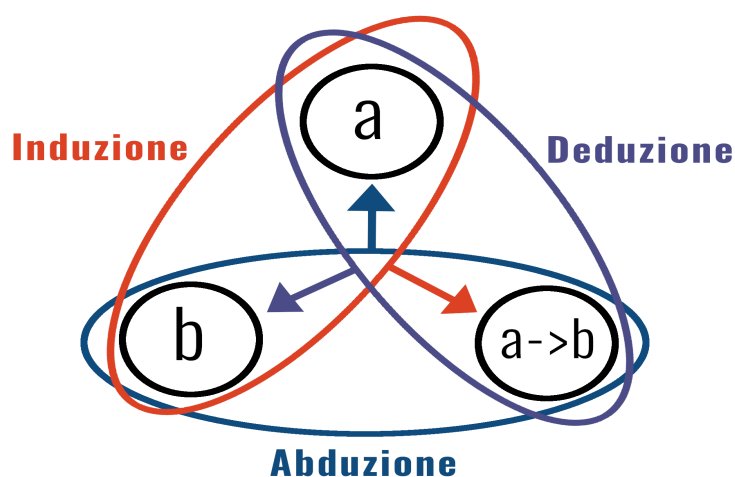


Figura 3

La coerenza di giustificatori indipendenti non è truth-conducive... ma il coerentismo lo nega: il teorema di Bayes

Nel coerentismo più sono i rapporti d'inferenza tra le credenze maggiore è il grado di giustificazione del sistema, e più il sistema è giustificato più le credenze che vi appartengono sono giustificate. Questo vuol dire che se una certa credenza è il giustificatore di un'altra, il giustificatore non ha bisogno di un'ulteriore giustificazione perché è già giustificato dalla sua coerenza con la credenza che giustifica. Ovviamente se ci sono credenze senza rapporti con le altre, il grado di giustificazione cala; e se ci sono credenze in contraddizione, cala di molto (cfr. Laurence Bonjour 1985).

Il sistema «Socrate è un uomo, è mortale, e se è uomo è mortale» è un sistema massimamente giustificato in quanto ogni credenza è in un qualche rapporto d'inferenza con le altre due.

Il sistema «Shakespeare è un grande scrittore, mi piacciono le carote, e un chilo di piombo pesa un chilo» non contiene contraddizioni, ma neanche prevede rapporti tra le credenze: non è un sistema coerente, è solo *non-incoerente*, non può essere detto giustificato.

Il sistema «Alberto è più alto di Barbara, Barbara è più alta di Carlo, Carlo è più alto di Alberto» è altamente incoerente. Le tre credenze non sono di per sé contraddittorie, ma se le prendo a coppie e calcolo le inferenze queste, invece di supportare le altre credenze del sistema, le negano:

- Alberto è più alto di Barbara, Barbara è più alta di Carlo: quindi Alberto è più alto di Carlo
- Alberto è più alto di Barbara, Carlo è più alto di Alberto: quindi Carlo è più alto di Barbara
- Barbara è più alta di Carlo, Carlo è più alto di Alberto: quindi Barbara è più alta di Alberto

È il tipo di sistema più incoerente: non solo le tre credenze non hanno rapporti inferenziali tra di loro, ma anzi ogni credenza è contraddetta dalle altre due (Thagard 2000).

Già discutendo i casi di Gettier, però, ho detto che la conoscenza non discende sempre da ragionamenti, e quindi il rapporto di giustificazione non può essere ridotto ai tre tipi d'inferenza visti: questo è essenziale non solo per il fondazionalismo, che *deve* avere giustificazioni non-inferenziali, ma anche per il coerentismo.

Miss Thagard invita nella sua camera da letto Jack lo Squartatore, tipo dal nome raccomandabile; un'ora dopo viene trovata morta dalla domestica. Io so che:

- Jack è stato fermato appena fuori dalla casa dei Thagard con un coltello insanguinato;
- era già noto a Scotland Yard in quanto aveva dei precedenti per spaccio;
- pochi giorni prima aveva postato sul suo account Twitter messaggi d'incitamento all'odio verso le donne e aveva minacciato di morte la stessa Miss Thagard.

Questo sistema è un insieme di prove che io raccolgo per giustificare la credenza che «è stato Jack ad ucciderla» (cfr. Olsson 2017). Sembra proprio essere coerente, molto coerente. Epperò c'è un problema: nessuna delle credenze è inferibile dalle altre due, e invero nessuna coppia di queste credenze può portare a concludere la colpevolezza di Jack neanche come abduzione. Se il sistema è coerente come sembra, e in generale può giustificare la mia credenza, deve esserlo in virtù di altro.

Mi chiedo: qual è la caratteristica essenziale della giustificazione? La giustificazione deve essere *truth-conducive*, deve condurre alla verità. Se so cos'è la verità dovrei anche sapere come fare a raggiungerla, e se so questo so anche cosa può essere un giustificatore: un giustificatore è ciò che aumenta la probabilità di una certa credenza di essere vera; la credenza «l'orologio segna le due» è in rapporto di giustificazione con «sono le due» se aumenta la probabilità che siano davvero le due. Questa definizione di giustificazione è *neutra*, cioè può essere adottata da qualsiasi teoria.

Riformulo quindi la teoria coerentista della giustificazione così: «una credenza è giustificata se aumenta la probabilità che le credenze di un sistema siano vere (o se le credenze del sistema aumentano la probabilità che la tal credenza sia vera) ed è tanto più giustificata tanto più le credenze del sistema aumentano *reciprocamente* la loro probabilità di essere vere» (cfr. Lewis 1946).

La teoria, così espressa, è coerente con l'indefettità. Per esempio, il giustificatore della credenza «sono le due» è l'orologio, cioè «se l'orologio segna le due, sono le due», e il suo metagiustificatore, cioè la credenza «l'orologio funziona tramite un meccanismo tale da seguire l'ora» aumenta la probabilità che la credenza sia vera; similmente, l'assenza di elementi che mettono in dubbio

l'affidabilità della credenza aumenta la sua probabilità di essere vera, e quindi il grado nel quale aumenterà la probabilità delle altre credenze; ma non tutto ciò che io so essere implicato da una conoscenza aumenta la probabilità del sistema, quindi non tutte le implicazioni sono giustificate e quindi il principio di chiusura non regge.

Il fondazionalista però può sollevare una forte obiezione. Vado a vedere un processo, l'accusa è «Aldo ha rubato». Viene chiamato a testimoniare Bruno, e dice che è vero, lo ha visto coi suoi occhi; poi viene chiamata a testimoniare Carlotta, e dice lo stesso. Tra Bruno e Carlotta c'è totale indipendenza, la testimonianza di uno non cambia in funzione dell'altro. Se io lancio una moneta, la probabilità che esca testa è sempre 50%, non importa quale sia stato il risultato dei lanci precedenti, appunto perché ogni lancio è indipendente. Similmente, l'ordine delle testimonianze è irrilevante alle loro verità, sarebbe controintuitivo dire che la probabilità della verità della testimonianza di Carlotta varia in base all'ordine col quale si chiamano i testimoni (per esempio: se Carlotta avesse testimoniato per prima avrebbe avuto una probabilità del 50%, ma poiché è seconda in virtù del suo accordo con Bruno ha un 60%). Ma se la probabilità della verità di una certa testimonianza è indipendente dalle altre, non importa quante siano e quanto siano d'accordo, il grado di coerenza dell'insieme delle testimonianze è sempre lo stesso. Eppure, più sono le testimonianze contro Aldo, più è probabile che sia colpevole. Quindi la coerenza del sistema non è *truth-conducive*, che le credenze a sistema aumentino reciprocamente la loro probabilità di essere vere è inessenziale alla giustificazione.

Posso analizzare matematicamente questo contrasto tra fondazionalismo e coerentismo. L'impianto generale dell'argomentazione seguente viene da Olsson (2002) e da Huemer (2011), quest'ultimo un filosofo magro come la morte, sosia di Keanu Reeves, e col quale sono quasi sempre molto d'accordo.

Thomas Bayes è stato un ministro presbiteriano e matematico del Settecento, forse il più grande statistico di sempre assieme a Ronald Fisher. È lo scopritore di un fortunatissimo teorema la cui importanza all'interno dell'epistemologia è in ascesa, tanto che gli ultimi decenni hanno visto lo sviluppo di un settore a parte, l'epistemologia *bayesiana*. Tale teorema permette di rappresentare il modo in cui la probabilità di una ipotesi varia man mano che nuove informazioni si rendono disponibili (per esempio come aumenta la mia giustificazione nel credere l'imputato colpevole man mano che ascolto nuove testimonianze). Questo è il *teorema di Bayes*:

$$P(B) = \frac{P(A)*P(A)}{P(B)}$$

$P(A|B)$ indica la probabilità della verità di un fatto A dato il giustificatore B, quindi $P(B|A)$ la probabilità della verità del giustificatore B dato il fatto A. Faccio un esempio, voglio calcolare la probabilità che il sole sorga ogni giorno della settimana dato il fatto che è sorto oggi:

- $P(A)$, cioè la stima iniziale della probabilità che il sole sorga tutti i giorni della settimana, la pongo arbitrariamente all'1%;
- $P(B)$, la probabilità che il sole sorgesse oggi cioè di lunedì, la pongo a $\frac{1}{7}$, quindi circa 14%;
- $P(A)$, la probabilità che il sole sorgesse oggi dando per vero che il sole sorge tutti i giorni, è ovviamente del 100%.

Quindi: $\frac{100*1}{14}$, circa 6,6%, il fatto che sia sorto oggi ha aumentato la probabilità della folle ipotesi che il sole sorga tutti i giorni.

Ora, nel coerentismo le credenze non possono essere giustificate in isolamento, un singolo giustificatore non aumenta la probabilità di una credenza. Penso di nuovo al processo: se tre testimoni accusano tutti l'imputato di furto, il loro accordo va spiegato con la verità del fatto e questo giustifica la sentenza, ma una singola testimonianza non giustifica perché non richiede la verità del fatto come spiegazione; posti A, B e C, né B né C giustificano A, ma «B e C» potrebbe (cfr. Bonjour 1985). Lo chiamo *principio di coerenza*:

$$\begin{aligned} P(B) &= P(A) \\ P(C) &= P(A) \\ P(B, C) &> P(A) \end{aligned}$$

Il fondazionalista però ha fatto notare che le varie testimonianze sono indipendenti. Se io so già che Aldo ha rubato, e Carlotta lo testimonia, la probabilità della verità della sua testimonianza è del 100%. E se assieme a Carlotta ci sono altri mille testimoni che dicono tutti il contrario di lei? Non cambia nulla, è sempre 100%, perché la verità della testimonianza dipende solo dal suo corrispondere alle azioni di Aldo. Questo *principio d'indipendenza condizionale* lo scrivo così:

$$P(A, B) = P(C|A)$$

Posto ciò, calcolo la probabilità di A dati sia B che C:

$$P(B, C) = \frac{P(A, B) * P(A)}{P(C)} = \frac{P(A) * P(A)}{P(C)} = \frac{P(C) * P(A)}{P(C)} = P(A)$$

Per il principio d'indipendenza la probabilità di C dati A e il giustificatore B è uguale alla probabilità di C dato il solo A, ma per il principio di coerenza la probabilità di C dato A è uguale alla probabilità di C; da ciò segue che, non importa quanti giustificatori si possano mettere insieme, la probabilità di A non aumenterà. Ma questo è assurdo.

I due principi non sono propriamente in contrasto, ma se combinati portano a conclusioni razionalmente inaccettabili: se si accetta il principio di coerenza non si può accettare quello d'indipendenza. Epperò rifiutare l'indipendenza condizionale è un azzardo: la sua correttezza è intuitivamente chiara, e l'incredibile fecondità della sua applicazione all'intelligenza artificiale non permette di scartarla a cuor leggero (e.g. Pearl & Azaria 1986). La soluzione più semplice è rifiutare il principio di coerenza, e accettare che anche una sola testimonianza ha potere giustificante.

Questo porta a rifiutare il coerentismo classico, il quale però ha comunque offerto varie intuizioni da conservare: l'idea che la coerenza del sistema possa giustificare *da zero* senza altri fattori a contribuire è sbagliata (Van Cleve 2011), ma che i giustificatori possano influenzarsi a vicenda è probabilmente vero, e così anche che nessun giustificatore possa rendere *razionalmente sostenibile* una credenza da solo. Credo inoltre che il coerentista non intenda la probabilità come qualcosa di totalmente riducibile alla statistica matematica (Thagard 2005), il senso comune viene prima della statistica, e anche l'esempio fatto col teorema di Bayes può sembrare più o meno convincenti a seconda dei presupposti di sfondo che portano ad assegnare le probabilità di partenza.

La tesi fondazionalista: quali credenze sono basiche?

Ho detto che una credenza basica è giustificata in modo non-inferenziale. Com'è giustificata, dunque? Se ogni inferenza richiede delle premesse, come sono individuate queste premesse che precedono ogni successivo ragionamento?

Cartesio è uno dei padri di questa scuola: dopo di lui il fondazionalismo si impone come la posizione standard, il coerentismo è un'alternativa recente. Nella sua filosofia le credenze basiche sono quelle *infallibili*, cioè quelle che se negate generano un paradosso, come «io penso», o che sono rese vere dal solo fatto di essere credute, come «io ho delle credenze» (Cartesio 1641).

Cartesio usa in modo molto liberale tali credenze: dal fatto di esistere deduce in fretta l'esistenza di Dio. Ad essere sinceri, tali credenze indubitabili sono piuttosto inerti: «penso, dunque sono», occhi, e allora? Non posso logicamente dedurre altro, mi fermo qui. Invero, non posso accettare che le credenze basiche siano solo le credenze indubitabili: serve qualcosa di più per poter ragionare e giustificare ciò che credo di sapere sul mondo.

Come minimo serve un insieme di credenze sulla logica. «Io non sto pensando» è paradossale perché è come dire «sto pensando, e non sto pensando». Bene: perché una proposizione di questo tipo non è accettabile? Ci sono logiche, dette *paraconsistenti*, nelle quali queste formule contraddittorie sono perfettamente gestibili (Priest 1998, Berto 2010). Io posso dire che «sto pensando» è indubitabile, e posso sperare di inferirvi altre credenze, solo se ho già una serie di convinzioni sulla logica, se ho un sistema che mi permette di dedurre proposizioni e giudicare la loro consistenza e accettabilità. Anche queste sembrano dover essere credenze basiche, è improbabile che siano giustificate da altro, cioè che siano la conclusione di un ragionamento, visto che sono un presupposto di ogni altro ragionamento (ma per questo *problema dell'adozione* cfr. Nicolai & Cohnitz inedito).

Anche le credenze derivate da percezioni, altro esempio di possibili credenze basiche, non sono autonome: io posso credere «qui c'è una mela» in virtù di una certa esperienza visiva solo se ho già gli schemi mentali che mi permettono di vedere una mela anziché una chiazza di colore. Perché davanti ad una sedia dovrei vedere una sedia, e non uno strano tavolino, o un divano molto alto per persone molto piccole? Certe illusioni ottiche che fanno impazzire gli psicologi mostrano come la visione degli oggetti formi un tutt'uno con le aspettative (Foss 1968), per esempio il *cuo di Necker* può essere interpretato come visto dall'altro o visto dal basso a seconda del modo in cui si processa l'immagine:

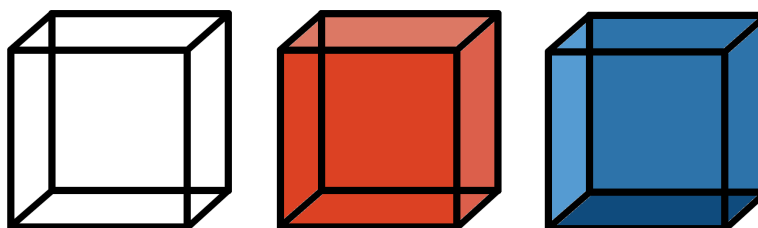


Figura 4

Se io nascessi come una *tabula rasa*, senza avere già nel mio cervello uno *spettro del mondo*, cioè senza una implicita aspettativa su ciò che sarà la vita, allora non potrei neppure sviluppare alcuna percezione. Queste aspettative costituiscono credenze? È discutibile, ma penso possano essere

espresse come tali: per esempio, il modo in cui cambia la visione del cubo di Necker può essere spiegata tramite la *credenza* su quale sia la faccia più profonda. Sono anche credenze basiche? Non è necessario ma possibile: la percezione le presuppone, credere che possano essere giustificate da credenze acquisite tramite le percezioni sarebbe circolare, né mi sembrano poter essere giustificate da altre delle credenze basiche che ho ipotizzato finora.

Continuando su questa linea, Chomsky sostiene un *innatismo linguistico*: l'estrema rapidità con la quale un bambino può apprendere una lingua e l'esistenza di caratteristiche universali in tutte le lingue del mondo suggeriscono che l'essere umano non scopre la lingua nascendo, ma già prima della nascita il cervello è predisposto ad apprendere un linguaggio, si deve *sintonizzare* col mondo nei primi anni di vita ma ha già tutti gli strumenti per fare ciò che deve fare (Chomsky 1975, Pinker 1998).

Queste non sono credenze, sono solo *disposizioni*; sono un po' come dei riflessi: quando metto una mano sul fuoco l'arco riflesso me la fa ritrarre, ma non c'è credenza, non credo di dover ritrarre la mano per far cessare il dolore. Tali disposizioni, però, potrebbero essere i giustificatori di certe credenze presupposte dal linguaggio (credenze elementari sul mondo come l'esistenza di quantità, credenze sulla possibilità di comunicare, eccetera), credenze che quindi risulterebbero basiche.

Certo, che una certa credenza sia innata, che preceda temporalmente tutte le altre, non significa che sia basiche: forse posso dire che un neonato nasce con certe credenze su come avere il latte, ma, se sono credenze, non sono giustificate in sé, saranno giustificate dalla prova empirica. Epperò il caso del linguaggio è diverso.

Nella filosofia analitica una credenza è un atteggiamento proposizionale (Russell 2003), un modo nel quale mi pongo nei confronti di una certa proposizione, e le proposizioni sono espresse dal linguaggio: quindi un animale che non ha un linguaggio non può neanche avere credenze in senso proprio (Wittgenstein 2014, Davidson 1994 e 2003). Il linguaggio umano pare presupporre una serie di credenze, le quali però non sono esprimibili a prescindere dal linguaggio: cioè sembra che, nel momento in cui acquisisco la capacità del linguaggio, e quindi nel momento in cui acquisisco la possibilità di avere credenze, in virtù del modo in cui il linguaggio è, acquisisco simultaneamente una serie di credenze (e forse *meta-credenze*, credenze sulle credenze) che sarebbe circolare giustificare con qualsiasi altra credenza. È un'idea che mi pare suggestiva, non so quanto fondata, e che potrebbe portare ad individuare ulteriori credenze basiche.

Considero infine le credenze *assiologiche*, cioè le credenze sul valore, come quelle etiche ed estetiche. Come detto, in Peirce il valore precede la logica, la verità è il valore delle proposizioni. Se ha ragione, è difficile trovare delle credenze che possano giustificare le convinzioni assiologiche in modo non-circolare, quasi tutte sarebbero successive e quindi presupporrebbero credenze sul valore. In etica mostrerò che ci sono filosofi che sostengono ciò, che certi aspetti della conoscenza presuppongono il valore. Non è allora un'idea peregrina considerare il valore come qualcosa di fondativo che, se è giustificato, lo è in modo non-inferenziale.

Quindi come possibili credenze basiche ho ipotizzato:

- credenze indubitabili e infallibili, accompagnate da credenze sulla logica;
- credenze derivate da percezioni, accompagnate dalle credenze di quello che ho chiamato *spettro del mondo*;
- credenze legate al linguaggio;
- credenze assiologiche.

Non credo certo siano tutti i tipi di credenza presentabili come basici. Non credo neanche che tutti questi siano davvero tipi di credenze basiche, sono sicuramente discutibili. L'idea è solo quella di una tassonomia della giustificazione, una classificazione di vari tipi di credenza sulla base dei modi nei quali possono essere giustificate. Ma allora: che tipo di credenza è la mia credenza nel fondazionalismo, e com'è giustificata?

Il fondazionalismo nega l'apertura epistemica: circolarità

Io vedo una mela. Come lo so, come mi giustifico? Suppongo sia una credenza basica: sono giustificato dalla visione la quale, non essendo a sua volta una credenza, non necessita di ulteriore giustificazione. Ma arriva lo scettico e mi chiede: «come lo sai che è basica? Come fai a sapere che è giustificata da una percezione, e non da un'altra credenza com'è in tutti gli altri casi? Come fai a sapere che la percezione può giustificare, e per di più senza essere a sua volta giustificata? Non dovresti giustificarla controllando la sua affidabilità?» Io ora posso difendere la posizione in qualsiasi modo voglia, posso dire che è così perché tale è il significato che dò alla parola «giustificazione», posso apportare qualsiasi giustificatore, e lo scettico comunque mi chiederà «perché mai questo dovrebbe essere un giustificatore della tal credenza? Come lo sai?»

Le domande dello scettico meritano risposta: infatti, è desiderabile che io sia in grado di difendere razionalmente l'adozione del fondazionalismo. Il mio rispondergli implica che sto accettando una componente internalistica nella mia teoria, cioè sto accettando di dover avere accesso al giustificatore e al suo metagiustificatore. Affinché il guardare l'orologio giustifichi la mia credenza sull'ora io devo sapere perché lo fa, devo sapere che funziona in un certo modo, e se non dovesse funzionare la mia giustificazione verrebbe invalidata. È vero che anche una percezione, come giustificatore, ha un metagiustificatore? Se io prendo un medicinale i cui effetti collaterali sono stati alterati di coscienza, difficoltà nell'introspezione, credenze paradossali, il mio vedere una mela non sarebbe sufficiente a giustificare il mio credere di vederla, dovrei anche sapere che il medicinale non sta dando quegli effetti; quindi il medicinale potrebbe invalidare l'uso della percezione come giustificatore, quindi anche la percezione deve avere un qualche metagiustificatore che il medicinale potrebbe rendere falso.

Dunque, io devo sapere che una credenza basica è giustificabile nel modo non-inferenziale in cui è giustificata, devo dimostrare che la verità della credenza sulla percezione è sufficiente alla sua giustificazione, ovvero io devo dimostrare il fondazionalismo stesso. Ma come posso farlo?

– « x è giustificato sse basico o inferito da una credenza giustificata» è giustificato sse y .

Come nella teoria della verità, i due «giustificato» devono stare su livelli diversi, ma non possono avere due significati diversi, come se vi fossero due parole nascoste in «giustificazione»: infatti qualsiasi diversa giustificazione io possa portare, questa continuerebbe a richiederne un'altra; è invece necessario che la mia teoria sia in grado di autogiustificarsi. Stando così le cose, ci sono due possibilità: il fondazionalismo stesso è giustificato o perché è basico, o perché inferito da una credenza giustificata.

Che il fondazionalismo sia basico vuol dire che l'idea che la credenza basica sia giustificata è a sua volta una credenza basica. Cosa la giustifica? Appunto il suo essere basica, si sta autogiustificando, ma il fondazionalismo non accetta giustificazioni circolari. Devo allora sostenere che il fondazionalismo è giustificato da una credenza giustificata, cioè devo immaginare che esista una

credenza basica, giustificata da una percezione, che giustifica delle credenze, le quali giustificano che la credenza basica è giustificata. Ma, accettando l'internalismo, la credenza basica non può essere (pienamente) giustificata solo dalla percezione, devo anche sapere che può essere giustificata così, e questo io lo so solo quando le credenze (debolmente) giustificate giustificano il fondazionalismo, di modo che la credenza basica finisce comunque per autogiustificarsi.

Questo segna un contrasto tra la mia concezione della giustificazione e il fondazionalismo: se quest'ultimo accettasse la necessità di una componente internalistica del tipo proposto, dovrebbe accettare anche i ragionamenti circolari, ma non può farlo. Inoltre, anche se si superasse questo punto, il fondazionalismo rimarrebbe incompatibile con il mio sistema in quanto resta una teoria cartesiana, che implica il principio di chiusura.

Discutendo lo scetticismo, io ho voluto rifiutare l'idea che la conoscenza sia *chiusa per inferenza*, ma non ho mai negato che la verità lo sia; cioè, io accetto la tautologia per la quale

$$\mathbf{T} \quad (p \wedge (p \rightarrow q)) \rightarrow q$$

Il fondazionalismo può essere inteso come l'idea che la giustificazione sia chiusa per inferenza, cioè che

$$\mathbf{F} \quad (Jp \wedge (p \rightarrow q)) \rightarrow Jq$$

Il problema è che se sia la verità che la giustificazione sono chiuse per inferenza, anche la conoscenza lo è. Infatti, suppongo di conoscere un fatto p , di sapere che implica q , e di credere q :

$$\mathbf{P1} \quad Kp \wedge K(p \rightarrow q) \wedge Bq$$

Poiché la conoscenza è credenza vera giustificata, le due proposizioni conosciute sono vere e giustificate, cioè P1 implica le proposizioni seguenti:

$$\mathbf{P2} \quad p \wedge (p \rightarrow q)$$

$$\mathbf{P3} \quad Jp \wedge (p \rightarrow q)$$

Ne consegue, per T ed F, che

$$\mathbf{P4} \quad (Kp \wedge K(p \rightarrow q) \wedge Bq) \rightarrow (q \wedge Bq \wedge Jq)$$

Ma, per la definizione di conoscenza,

$$\mathbf{P5} \quad (q \wedge Bq \wedge Jq) = Kq$$

e così ottengo il principio di chiusura della conoscenza. Se non voglio farlo sorgere, devo rifiutare uno dei presupposti, e l'unico di questi ad essere rifiutabile è F, cioè il fondazionalismo stesso.

Detto tra parentesi: il positismo

Non potendo io accettare un fondazionalismo del tutto esternalista, se questo continua a sostenere l'inaccettabilità di ragionamenti circolari, le credenze basiche risultano semplicemente ingiustificate. Ebbene, questo può non essere un gran problema, secondo Wittgenstein.

Wittgenstein conosceva molti dei più grandi filosofi della sua epoca. Tutti i più grandi, quasi tutti affiliati all'università di Cambridge, passeggiavano e discutevano lungo le stesse strade: Wittgenstein, Popper, Russell, Moore, Ramsey, il grande economista Keynes, si frequentavano tutti. E Wittgenstein era a tutti noto come un conversatore tanto stimolante quanto irritante. Quando scrisse all'economista italiano Sraffa implorandolo di restare suo amico e promettendogli che avrebbero parlato di tutto ciò che lui avrebbe voluto, Sraffa rispose «parleremmo di ciò che voglio io, ma solo nel modo in cui vuoi tu» (peccato, visto che era proprio Sraffa lo stimolo di alcune delle più geniali intuizioni delle *Ricerche Filosofiche*) (Monk 2000). Tra gli amici di Wittgenstein rimasti vittima della sua furia oratoria vi era anche Popper.

Una volta Wittgenstein stava presentando una ricerca nella quale tentava di dimostrare che non esistono problemi filosofici, ogni problema filosofico è solo una confusione di linguaggio, è in realtà un problema *lessicografico*. Popper non era d'accordo, credeva nell'esistenza di genuini problemi filosofici. La discussione si scaldò a tal punto che Wittgenstein prese un attizzatoio e lo fece vibrare a destra e a sinistra per sottolineare le sue affermazioni; quando Wittgenstein gli chiese «fammi un esempio di regola morale» Popper rispose «non minacciare gli ospiti con l'attizzatoio», al che Wittgenstein lo gettò a terra e se ne andò (Edmonds & Eldinow 2005).

All'epoca, Popper aveva già rivoluzionato la filosofia della scienza. Vado a Venezia per la prima volta, vedo una gondola nera e dico «a Venezia, tutte le gondole sono nere». L'amico filosofo mi risponde «eh no, tutto ciò che possiamo dire è che a Venezia c'è almeno una gondola nera, anzi, che c'è almeno una gondola che è nera dal lato che abbiamo osservato». In effetti, per quante gondole io possa osservare, non potrò mai confermare in modo certo la mia teoria «tutte le gondole sono nere», posso osservarne un milione di nere, questo non mi dimostra che non ne esistono di rosse; di contro, mi basterà osservarne solo una di colore diverso, e sarà sufficiente a dimostrare che la mia teoria è falsa. Tutte le volte che ho visto qualcuno calciare un pallone, ho visto che poi questo pallone volava via; ma non c'è proprio nulla che mi garantisca che la prossima volta non lo vedrò invece esplodere o passare attraverso il piede. Nessuna teoria empirica può essere dimostrata vera, può essere solo dimostrata falsa. Questo è il concetto che sta alla base del *falsificazionismo* di Popper (2010).

Popper non credeva invece nel *giustificazionismo*, l'idea che la teoria della giustificazione sia il cuore dell'epistemologia: secondo lui le si dà troppa importanza, è una visione assolutistica della conoscenza ancora dipendente dall'ideale utopico della certezza. Mentre sono sotto la doccia penso «forse l'elettrone è un buco nero che accelera all'indietro nel tempo...» Un giustificazionista non mi prenderà sul serio, magari ho pensato qualcosa di vero ma non ho alcuna ragione per crederlo, è un'idea che non merita attenzione. A Popper, invece, non interessa quali siano i miei giustificatori, gli interessa solo la possibilità di mettere alla prova quella credenza: la conoscenza non avanza in virtù della giustificazione, avanza escludendo le credenze che si dimostrano essere false, che queste fossero giustificate o frutto del caso è irrilevante, erano testabili e quindi meritevoli di attenzione (Popper 2010).

Anche se Wittgenstein non amava molto visione filosofica di Popper, il suo grande scritto epistemologico, *Sulla Certezza*, pare riconoscere i meriti di questo deflazionismo nei confronti della giustificazione.

Wittgenstein aveva uno stile filosofico peculiare, sia come scrittura che come modo di portare avanti la riflessione: vedeva la filosofia come una terapia, i problemi filosofici sono malattie del pensiero e del linguaggio, il filosofo-medico deve curarli per riportare ordine e quiete nella mente (W. 2014). Se notava un contrasto tra due posizioni intellettuali, Wittgenstein non cercava di fonderle prendendo il bene da entrambe, al contrario individuava i loro presupposti comuni per poi negarli e rifiutarle entrambe in un colpo solo. Wittgenstein riconosce che sia lo scetticismo sia il fondazionalismo partono dal presupposto che solo una credenza giustificata possa giustificarne un'altra. E questo è ciò che lui, implicitamente, nega: le credenze basiche devono giustificare tutte le altre, per questo non possono essere giustificate a loro volta (W.1999).

Questa teoria è detta *positismo*: le credenze basiche sono *posit* ingiustificati che si pongono proprio perché se non si ponessero sarebbe impossibile giustificare alcunché (Wright 2004). Da qualche parte bisogna pur partire!

Invero l'epistemologia dell'ultimo Wittgenstein è molto più complessa di così: ha anche una parte di pragmatismo (il mio credere o dubitare è rilevante solo quando porta a differenze di tipo pratico) e di coerentismo (i *posit* sono contenuti in sistemi di credenze che vanno giudicati nell'insieme) (W. 1999). La cifra della teoria però rimane l'idea che una credenza ingiustificata possa giustificare le altre. Questa è una possibile soluzione al problema della circolarità: un fondazionalista può accettare che le credenze basiche siano essenzialmente circolari, e che il ragionamento circolare non possa giustificare, senza per questo negare che le credenze basiche possano giustificare tutte le altre.

Ora devo chiedermi: quali sono questi *posit*? I *posit* sono «quelle credenze che, se negassero, la giustificazione sarebbe impossibile». Il *posit* più ovvio è «non sono in un contesto scettico» perché se lo negassi non potrei giustificare nulla. Ma poi? Non sembra possibile dire molto di più. Lo stesso Wittgenstein pare sposare una visione molto *continentale* per la quale i *posit* sono determinati da fattori culturali e storici. Questo mette in mostra tutta l'arbitrarietà della quale si può accusare un fondazionalismo.

Huemer (2001, 2007), come Wittgenstein, nega che le credenze legittime necessitino sempre di giustificazione, ma la sua teoria è ancora più estrema: per lui è una credenza legittima ogni credenza che *al momento* non ha prove contro, sono giustificato *di default* a credere ciò che credo fino a prova contraria. Accettare questa teoria può dare effetti indesiderati: è impossibile avere prove contro un contesto scettico, quindi, se mi sembrasse di star sognando, sarebbe sempre e *per sempre* legittimo credere di stare sognando.

Quindi nessuna delle due teorie è soddisfacente 1: il fondazionalismo aperto

Se neanche trattare le credenze basiche come *posit* funziona, il fondazionalismo, quello che io posso accettare almeno, non ha via di scampo: deve tollerare i ragionamenti circolari e accettare una nozione più larga di giustificazione che non si limiti all'inferenza tra credenze non-basiche. Credo se lo possa permettere: il fondazionalismo rimane tale fintanto che afferma l'indipendenza dei giustificatori, prevede credenze basiche giustificate in modo non-inferenziale, e accetta che anche una credenza isolata possa giustificarne un'altra.

Ma oggi i fondazionalisti non si fermano qui, sono pronti a fare ulteriori concessioni per sviluppare una teoria più vicina al senso comune. Immagino due sistemi di credenze:

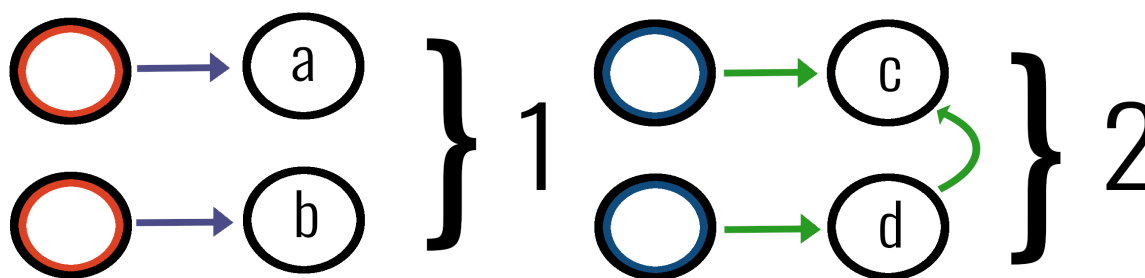


Figura 5

Mi chiedo: perché mai il secondo sistema non dovrebbe essere preferito al primo? Nel primo immagino di avere due osservazioni qualsiasi, dalle quali derivo le credenze A e B. Nel secondo pure derivo C e D da due osservazioni, ma poi scopro anche una teoria che da D deduce C. In questo caso C non solo è confermata dall'osservazione, ma è anche prevista dalla teoria, e in virtù di questa doppia giustificazione C è più giustificata di A e B; ma proprio perché C è così fortemente giustificata, mi dà motivi di credere che ciò da cui l'ho derivata sia corretto, cioè anche la probabilità della correttezza di D e dell'osservazione aumenta; così D, pur non essendo implicata da altro, come A e B, risulta comunque più giustificata di loro. Per di più, questo mi dà motivo di credere che i miei processi osservativi siano generalmente affidabili, giustificando anche altre future osservazioni. Cioè, se le credenze non-basiche C e D possono influenzarsi e rafforzare la loro giustificazione, tutto il sistema risulta più credibile.

Il fondazionalismo così accetta che le credenze possano formare un sistema nel quale parte della giustificazione è *olistica*. Questo avvicina molto il fondazionalismo al coerentismo.

Dall'altra sponda, i coerentisti pure stanno integrando delle caratteristiche del fondazionalismo, e hanno tutto l'interesse a farlo. L'accusa che Russell (1907) muove contro il coerentismo come teoria della verità ha degli effetti anche in teoria della giustificazione. È infatti intuitivo che esistano infiniti sistemi di credenze possibili con uno stesso grado di coerenza ma composti da credenze diverse e anche opposte. Uno può credere che la Terra giri attorno al Sole, oppure che la Terra sia il centro dell'universo, con pianeti che viaggiano lungo orbite a forma di fiore e stelle e galassie che vibrano nel cielo. Entrambi questi sistemi sono coerenti, e la coerenza può essere davvero *truth-conducive* in questi casi, ma lo è in modo debole e insufficiente perché nulla mi dice su quale sistema dovrei preferire (cfr. Kuhn 2009). Questo può essere accettabile in matematica (per esempio esistono più geometrie, più insiemi coerenti di assiomi e teoremi), ma non lo è se si vuole arrivare ad un ritratto della realtà. Per questo i coerentisti ora accettano l'esistenza di virtù di un sistema che vanno al di là della sua coerenza; è cosa comune, inoltre, accettare che certe credenze, come quelle derivate da esperienze, abbiano un ruolo speciale, diano anche in isolamento un maggior grado di giustificazione e rendano il sistema che le comprende più desiderabile.

Così, fondazionalismo e coerentismo si avvicinano sempre di più fino a collapsare in un'unica teoria (Haack 1993). Quando mi chiedo se una certa credenza è giustificata, se uso una prospettiva fondazionalista analizzerò le caratteristiche di quella credenza per capire se è basilica e quali sono i suoi rapporti inferenziali; se uso una prospettiva coerentista invece non la guarderò neanche perché devo analizzare le caratteristiche del sistema in cui è inserita. Questi due approcci non sono in contraddizione e non sono alternativi, sono entrambi aspetti della giustificazione, e vanno portati avanti in contemporanea.

Thagard (2000) ha sviluppato un coerentismo strutturato in sette principi che inglobano aspetti di entrambe le teorie. Questo che segue è una sua versione semplificata (cfr. anche Bonjour 1985):

- più un sistema di credenze è coerente, più è accettabile (questa è la parte coerentista);
- credenze derivate da osservazioni sono accettabili *di per sé* (questa è una parte fondazionalista);
- maggiore è il numero di credenze non inferite da altre (quindi ingiustificate), minore è la coerenza;
- siano A, B, C tre credenze qualsiasi, tali che: da A si inferisce C, e da B si inferisce C;
 - A è coerente con C, e B è coerente con C, ognuno allo stesso modo;
 - se da A si può inferire B o viceversa, e da entrambi si inferisce C, allora A e B sono coerenti;
 - se invece A e B non sono in alcun rapporto inferenziale, e da entrambi si inferisce C, sono incoerenti;
 - A e non-A sono tra loro incoerenti; ogni contraddizione è composta da parti incoerenti.

Voglio proporre un abbozzo di teoria simile a questa. Il sistema che espongo lo chiamo *fondazionalismo aperto*.

Distinguo due livelli di giustificazione: quello individuale e quello olistico. Una proposizione è giustificata in modo individuale quando è indefettata, e quindi anche quando è infallibile o prodotta indefettatamente dall'osservazione; è invece giustificata in modo olistico, anche senza essere individualmente giustificata, in virtù della sua appartenenza ad un sistema con certe caratteristiche virtuose (cioè coerente in senso generale). Da una proposizione-giustificatore (che fa coppia con una proposizione-metagiustificatore) posso inferirne un'altra, e così facendo trasmetto la giustificazione: se una proposizione è individualmente giustificata, anche le proposizioni che la inferiscono o che sono da essa inferite, e che non sarebbero di per sé individualmente giustificate, ricevono una percentuale variabile della sua giustificazione individuale.

Per chiarezza, posso rappresentare il grado di giustificazione di una credenza così:

	Proposizione A	Proposizione B	Proposizione C
GI: Giustificazione individuale	X	Y	Z
GO: Giustificazione olistica	W	W	W
GT: Gius. totale (gius. se > GI e GO)	XW	YW	ZW

Ho voluto calcolare la giustificazione totale come una moltiplicazione per sottolineare che una proposizione, per essere giustificata *tout court*, deve avere un qualche grado di giustificazione ad entrambi i livelli, nessuno dei due può essere zero; anzi, dato che GT deve avere un valore superiore a GI e GO, entrambe devono avere un certo alto grado di giustificazione (superiore a uno), non possono essere troppo basse.

Questa necessità di un doppio livello di giustificazione può essere controintuitiva: se una proposizione è infallibile, perché ciò non basta? Una proposizione è infallibile, e per questo autogiustificata, se il negarla implica una contraddizione. Ma se sia così, dipende dalla logica adottata. Ci sono infinite logiche possibili, alcune talmente inconsistenti da derivare contraddizioni

da ogni negazione: così, relativamente a certe logiche, ogni proposizione ha un grado di giustificazione. Certo io non credo a tutte le proposizioni, ma la giustificazione è indipendente dalla credenza, ci sono proposizioni che sono giustificate a credere ma che non credo. Sono dunque giustificato a crederle tutte? Ovviamente no. E perché no? Perché, lo facessi, crederei in un sistema pieno di contraddizioni, cioè le proposizioni avrebbero una GO troppo bassa.

Per essere giustificato non basta che le proposizioni siano infallibili, serve anche che siano prodotte in modo coerente, cioè che discendano da una logica *creduta* e avente caratteristiche desiderabili. Questa logica sarà inevitabilmente un posit: non ho motivi di favorirla ad altre logiche, ma non posso avere motivi prima di adottare una logica, quindi la credenza nelle sue regole d'inferenza è da considerarsi giustificata in via eccezionale. Dirò dunque che il grado di GI rappresenta la struttura inferenziale che la credenza ha in questa logica, mentre GO indica quanto il sistema logico al quale appartiene la credenza in questione si avvicina al sistema logico ideale per la giustificazione, cioè un sistema tale che:

1. parte dal minor numero di premesse possibile...
2. ...premesse che sono, nella più alta percentuale possibile, proposizioni infallibili o prodotte indefettatamente dall'osservazione;
3. inferisce da questo piccolo insieme di partenza un gran numero di proposizioni, coprendo tutti gli àmbiti di interesse;
4. ha un rapporto tra il numero di proposizioni e numero di legami inferenziali tra di esse il più possibile vicino a 0 (cioè più legami che proposizioni);
5. non contiene contraddizioni, né proposizioni in competizione;
6. rispetta le *clausole anti-gettier*: le proposizioni sono, nella più alta percentuale possibile, indefettate (cioè non devono esservi elementi che mettono in dubbio la validità dell'inferenza, cioè devono darsi i fatti indicati dal metagiustificatore).

In questo modo, non c'è vera GT per inferenza e posso accettare l'apertura epistemica. L'idea è che la GI cali man mano che mi allontano dalla fonte, cioè che le conoscenze trasmettano solo una percentuale di giustificazione alle inferenze. Nello specifico, il valore di GI massimo trasmissibile tramite inferenze è 0,9, quindi se la credenza inferita non è essa stessa indefettata non avrà mai un GT sufficiente. Posso inoltre accettare l'indipendenza condizionale in quanto anche una sola proposizione individualmente giustificata trasmette giustificazione per inferenza, a prescindere dagli altri elementi del sistema. Infine, non ci sono problemi di regressione infinita o di auto-justificazione del fondazionalismo aperto in quanto nulla vieta i ragionamenti circolari: è vero che anche le proposizioni infallibili od osservative, dovendo essere indefettate, devono avere un metagiustificatore giustificato, ma che questo metagiustificatore sia una delle loro implicazioni (o la stessa logica adottata) non è problematico, né è problematico che la inferisca proposizioni equivalenti a questa stessa teoria (o che la teoria si ponga da sola come posit).

Voglio evidenziare il secondo punto, il quale mostra la possibilità di un *pluralismo contestualista* della giustificazione. È corretto assegnare alle proposizioni legate all'esperienza un ruolo particolare, ma ciò non può essere applicato indiscriminatamente a tutti i campi del discorso: se si parla di fisica è comprensibile che i sistemi più giustificati siano quelli che più si basano sull'osservazione; se però si parla di matematica le osservazioni sono irrilevanti, le proposizioni matematiche sono giustificate solo in virtù della coerenza interna ed è ingiusto richiedere di partire da credenze osservative pena una diminuzione del grado di giustificazione. Penso allora ad un sistema che assegni ai vari punti un peso diverso, una diversa importanza a seconda di quale sia il campo del discorso in questione. Posso però anche rispondere che la teoria della giustificazione deve riguardare le credenze nel loro complesso, deve riguardare il sistema di tutti i sistemi, e non la si

può limitare ad un certo campo; in tal caso, ciò che la teoria dice, pur essendo un'affermazione molto irriverente, è che un sistema di proposizioni esclusivamente matematico è meno giustificato di uno esclusivamente fisico.

Anche il quinto punto deve essere approfondito. Cosa sono le proposizioni in competizione?

Quindi nessuna delle due teorie è soddisfacente 2: possibili gestioni del conflitto tra sistemi

Proposizioni in competizione sono quelle che offrono spiegazioni per uno stesso fenomeno senza essere tra loro in alcun rapporto. La fisica atmosferica mi spiega perché ci sono fulmini; ma anche la mitologia greca me lo spiega, è Zeus che lancia saette fiammeggianti. Queste due credenze non hanno alcun legame: sono spiegazioni alternative, come due teorie scientifiche incompatibili. Un sistema che le prevede entrambe si trascina dietro un contrasto insanabile che ne riduce la coerenza.

Thagard (2000) interpreta il rapporto di giustificazione anzitutto come un rapporto inferenziale, e il rapporto inferenziale anzitutto come un rapporto di spiegazione (Poston 2004), quindi proposizioni in conflitto sono quelle dalle quali, senza che abbiano rapporti inferenziali tra loro, si inferisce una stessa altra credenza. Certo si tratta di una interpretazione ristretta di «giustificazione», valida solo in alcuni casi: il guardare l'orologio non spiega in nessun senso rilevante il fatto che sono le due (né il fatto che sono le due spiega l'ora indicata dall'orologio), eppure mi giustifica. E il fatto che una certa credenza sia inferita da altre due senza rapporti inferenziali tra loro non indica sempre la presenza di una coppia in competizione: sia guardare l'orologio sia vedere i negozi abbassare le serrande giustificano la credenza sull'ora, ma non sono in competizione. Epperò è vero che la competizione può esserci. Come gestire quei casi?

Io ho un sistema di conoscenze scientifiche: con un piccolo insieme di credenze in gran parte osservative riesco a giustificare e spiegare un gran numero di fenomeni del mondo, tutti connessi tra loro in modo coerente come le parole di un cruciverba. Dei cinque punti proposti, l'unico che questo sistema non soddisfa è il terzo: ci sono ambiti di alto interesse le cui relative credenze non paiono derivabili da questi fatti naturali. Immagino che le proposizioni etiche siano tra queste, e immagino di scoprire che un gran numero di proposizioni etiche sia derivabile in modo coerente da un piccolo insieme di premesse teologiche individualmente giustificate. Nell'accettare questo nuovo insieme di premesse riduco la coerenza per i punti uno e due, ma la recupero per tre e quattro: il sistema che ne risulta è ugualmente coerente ma spiega molte cose in più, quindi è favorevole a quello solo fisico. Ma la cosa si complica: noto una competizione tra i due sottosistemi, un piccolo insieme di credenze, magari quelle legate all'origine dell'uomo, che riceve sia una spiegazione fisica, sia una spiegazione teologica.

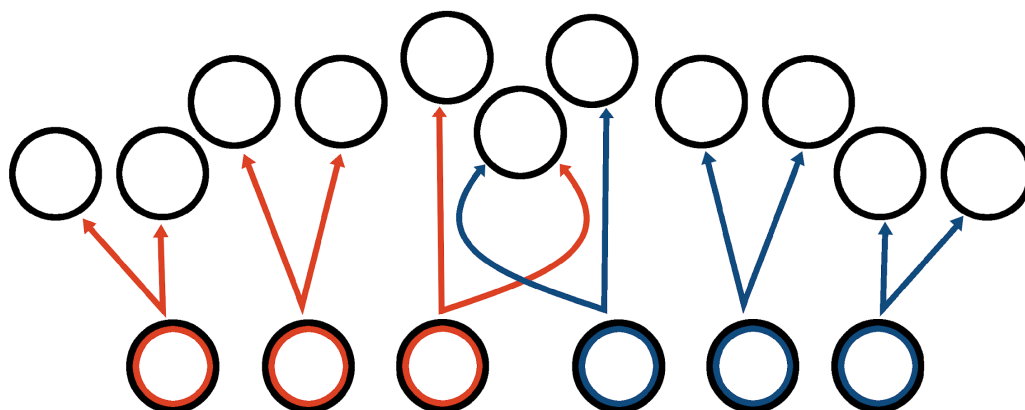


Figura 6

Escludere uno dei due sottosistemi è indesiderabile: è vero che è meglio non avere alcuna credenza etica piuttosto che avere credenze tirate a sorte; ma in questo caso il sottosistema teologico-etico è coerente, virtuoso e giustificato quasi tanto quanto quello fisico, ed è meglio avere tali credenze piuttosto che non averle. Bisogna quindi gestire il conflitto.

La prima possibilità è inserire un meccanismo di autocontrollo del sistema che, in caso di doppia spiegazione, blocca una delle due, salvando quella che proviene dal sottosistema con più alto grado di coerenza. Nell'esempio del conflitto tra scienza e teologia questa è la risposta standard, tutti (tranne i fanatici) accettano l'evoluzione come spiegazione dell'origine dell'uomo e considerano la Genesi come spiegazione di altri aspetti dell'esistenza. Il biologo Jay Gould (1997, 2002) ha accettato questa via: per lui scienza e religione sono *non-overlapping magisteria*, uno si occupa di fatti e l'altro di valori e non può esserci conflitto.

La seconda possibilità è sforzarsi di spostare uno dei due sottosistemi ad un livello più basso dell'altro. Continuando l'esempio di scienza e religione, i teologi con interessi scientifici tendono a seguire questa strada: considerano le credenze basiche della teologia come più elementari di quelle scientifiche, e cercano di mostrare che le credenze basiche della scienza sono derivabili e giustificabili da quelle della religione; oppure, al contrario, si sforzano di dimostrare che in qualche modo la scienza implica le credenze religiose. Così il conflitto scompare perché le due spiegazioni non sono più indipendenti.

La terza possibilità è accettare il conflitto. Questo è ciò che i filosofi del medioevo tendevano a fare: accettavano per fede che Dio ha creato il mondo e per ragione che il mondo è eterno, è un conflitto che riduce la coerenza ma è preferibile all'escludere uno dei due sistemi o al porre soluzioni *ad hoc* (d'Aquino 1270).

LETTURE PER APPROFONDIRE

Di Nozick e Drestke, i filosofi che rifiutano il principio di chiusura epistemica, si vedano anche Nozick 1999 e 2004 e Drestke 2000. Parte di ciò che credo in epistemologia è ispirato a Fraassen 1984, che sottolinea in modo secondo me corretto il ruolo della volontà nella credenza (senza uno slancio di volontà è impossibile pervenire a conoscenze, secondo me). Di Lehrer, invece, si

aggiungano 1974 e 1990. Per l'infinitismo Turri & Klein 2014. Di Huemer infine mi sono stati utili anche 1997 e 2007.

Sia Klein che Huemer sono influenzati da Paul Moser, il mio filosofo della religione preferito: si vedano 1985 e 1991. Per libri di altri pensatori su come gestire il possibile conflitto tra credenze religiose e credenze scientifiche si vedano almeno Polkinghorne 2000 e McGrath 2016.

Introduzione ai ragionamenti d'etica

L'importanza dell'etica

Quale può essere il valore del lavoro di un filosofo che se ne sta a rimuginare su una poltrona tutto il giorno?

La John Templeton Foundation è un'associazione filantropica interessata ai temi della moralità, dell'intelligenza, dello sviluppo personale, della creatività. Il Premio Templeton, assegnato a coloro che più hanno contribuito alla comprensione della religione e della spiritualità, consiste letteralmente in una montagna di soldi: nel momento in cui scrivo è pari a un milione cinquecentomila dollari, quasi il doppio di quel Premio Nobel che l'associazione accusa di trascurare i temi umanistici. Orbene, nel 2013 la Templeton Foundation ha posto due noti filosofi, John Greco ed Eleonore Stump, alla guida di un progetto triennale sull'umiltà intellettuale. La sovvenzione? Tre milioni trecentomila dollari.

Non è un'esagerazione. L'umiltà intellettuale è un valore etico, e le domande etiche sono le più importanti che l'umanità possa porsi. Mentre tutte le altre questioni filosofiche possono sembrare astratte e inutilissime, piante che non danno frutti, così non si può dire per l'etica, l'etica è il settore più pericoloso: dalle risposte che si danno discende la politica, per questo può avere conseguenze enormi e tragicamente concrete.

L'etica non parla solo di dilemmi e scelte difficili: parla di come si dovrebbe vivere, di cos'ha importanza, di com'è possibile creare scale di valori. Sono domande che mettono davanti alla propria abiezione, che stravolgono la visione del mondo e possono anche portare a crisi esistenziali: così è successo a me.

Nella filosofia analitica l'etica si divide in due settori principali: l'*etica normativa* e la *metaetica*. L'etica normativa si occupa della *forma* delle teorie etiche, cerca un criterio che permetta di compiere scelte di valore, ma non specifica quali siano queste scelte. La metaetica sta all'incrocio dell'etica con l'epistemologia e la metafisica, si chiede quale sia la natura del bene e come conoscerlo. Questi due discorsi sono abbastanza indipendenti: partirò con la parte normativa in modo da evidenziare poi come le diverse teorie reagiscono a diversi presupposti metafisici.

Non parlerò invece di etica descrittiva, il settore che analizza le effettive credenze morali di una popolazione dal punto di vista psicologico e sociologico; e neanche di etica applicata, quel settore che cerca di capire cosa fare in casi specifici e spesso d'interesse politico (i temi forti sono l'aborto e l'eutanasia, parti della *bioetica*, ma c'è anche un'*etica aziendale*, un'*etica militare*, eccetera).

Etica: etica normativa

Terzo ragionamento: proposta di un'etica deontologica

In questo capitolo sostengo una teoria etica personale: combino vari principi in modo tale, credo, da riuscire a rispondere a molti problemi diversi. Il ragionamento di base è il seguente:

- P1** il bene non può prevedibilmente diminuire la felicità né media né totale...
- P2** ...a meno che non sia richiesto dalla giustizia, dunque il bene non è riducibile alla felicità;
- C1** quindi l'utilitarismo è inaccettabile;
- P3** il bene è un dovere universalizzabile, sensibile alle motivazioni e sempre rispettabile;
- P4** il dovere è contestuale;
- P5** se infango un dovere, sono responsabile delle sue conseguenze (anche se non le controllo);
- P6** se supererogo, acquisisco nuovi doveri;
- P7** un dovere può avere effetti sia positivi che negativi purché indipendenti e inevitabili;
- P8** il dovere non può causare solo effetti negativi (neanche per impedire un male), ma può chiedere di non impedire l'effetto negativo di altre azioni;
- P9** il dovere può essere ad personam;
- P10** il dovere varia con le conseguenze intese;
- C2** la deontologia da sola non permette di variare il dovere con le conseguenze intese, quindi è inaccettabile;
- C3** l'etica della virtù permette di variare i doveri con le conseguenze intese, ma non dà una descrizione dei doveri che rispetti tutti i principi esposti, quindi è inaccettabile;
- C4** ma la combinazione di deontologia ed etica della virtù può rispettare tutti i principi esposti, quindi è accettabile.

Le tre tesi

L'etica valuta i comportamenti per decidere quale favorire in vari contesti. Il comportamento etico è una funzione a due posti: da una parte c'è l'azione, dall'altra lo stato futuro verso il quale il mondo evolverà come sua conseguenza. A seconda di dove si focalizza l'attenzione nella valutazione individuo tre grandi teorie:

- se valuto solo le conseguenze, uso un’*etica consequenzialista*;
- se valuto solo le caratteristiche dell’azione, uso un’*etica deontologica*;
- se valuto la funzione, i modi in cui si relazionano azioni e conseguenze, uso un’*etica della virtù*.

Per il consequenzialismo il bene sta in uno specifico stato futuro migliore di tutti gli altri, ma ci sono infiniti modi per arrivarci e ognuno è equivalente: valuta i fini e non i mezzi. Per la deontologia il bene sta in una specifica azione, c’è solo un modo d’agire rispettoso di tutti i diritti e i doveri, ma quali siano le sue conseguenze è irrilevante, ci sono infiniti futuri accettabili: valuta i mezzi e non i fini. Per l’etica della virtù il bene sta in principi-guida che, pur non valutando in modo particolare né mezzi né fini, li soppesano entrambi e accettano solo certe accoppiate.

Questa è una classificazione personale, tanto semplice quanto caricaturale: ogni eticista di buon senso, qualsiasi sia la teoria che sposa, tiene sempre in considerazione sia azioni che conseguenze.

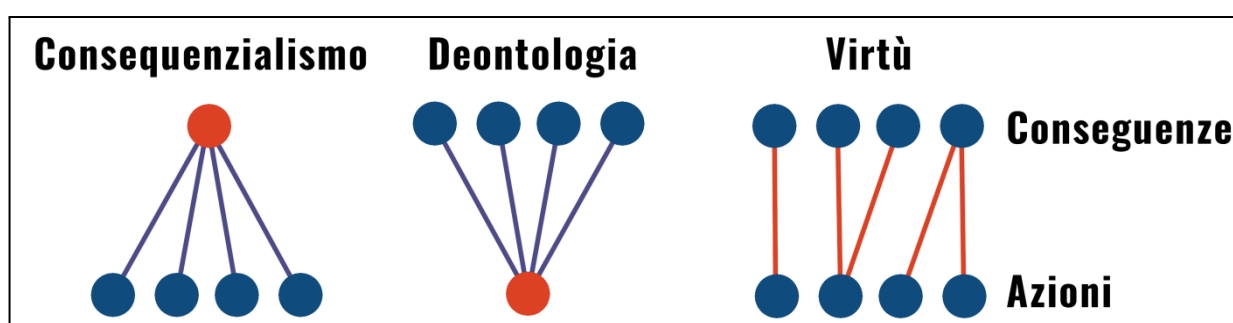


Figura 7

La tesi consequenzialista 1

Inizio col consequenzialismo: un’azione è buona tanto quanto lo è il futuro che realizza, più azioni che realizzano lo stesso futuro sono parimenti buone. Quindi devo chiedermi: quali caratteristiche mi permettono di distinguere un futuro desiderabile da uno indesiderabile?

La Storia ha proposto risposte diverse. Per i pensatori classici cinesi, con le loro preoccupazioni politiche, è buona l’azione che aumenta la ricchezza dello Stato, la sua popolazione, il suo ordine (questo è il *consequenzialismo statale del moismo*) (cfr. Johnston 2010). Per gli evoluzionisti filosofici un’azione è buona se aumenta le proprie probabilità di sopravvivere e di tramandare i geni, perché questo è ciò che la Natura ha premiato e premia (Dawkins 1995). Per gli anarcocapitalisti come Ayn Rand un’azione è buona se aumenta la propria libertà e ricchezza a prescindere dal benessere altrui (qui c’è *individualismo filosofico* ed *egoismo etico*) (Rand 1999, Stürner 1999). Al contrario, c’è chi sostiene un *altruismo radicale*, è bene ciò che aumenta il benessere altrui a prescindere dal proprio (cfr. Singer 2016).

La principale teoria consequenzialista, quella con la più storia più nobile e complessa, è l’*utilitarismo*. Nell’utilitarismo un’azione è bene se genera *utilità* (cosa essa sia dipende dall’utilitarista al quale lo si chiede).

Nonostante abbia una lunga tradizione che può risalire fin’anche a Platone o al grande oratore Demostene (1987), il fondatore dell’utilitarismo così com’è inteso oggi è Jeremy Bentham.

Bentham è stato un pensatore originale, radicale, e inquietante. È sua l'idea del *panopticon*, la prigione a pianta circolare che avrebbe permesso di osservare ogni prigioniero senza essere osservati: in questo modo i prigionieri, non potendo sapere se e quando li si sta osservando, sarebbero rimasti sempre sotto una pressione che li avrebbe spinti a comportarsi bene spontaneamente. Inoltre, i prigionieri avrebbero potuto passare il tempo correndo come criceti su una ruota che avrebbe alimentato meccanismi, come un mulino, per svolgere un lavoro utile che compensasse i costi della prigione (Bentham 1997). All'epoca questa tortura sembrava un modo per dare ai prigionieri migliori condizioni, dargli qualcosa di utile da fare invece di impazzire buttando tutto il loro tempo.

Un altro fatto per il quale è noto è l'*autoicona*. Era desiderio di Bentham che, dopo la sua morte, il suo corpo fosse preservato; così, una volta dissezionato, il suo scheletro è stato ripulito, vestito, messo seduto, e riempito di paglia, la sua testa pure è stata tagliata e conservata mummificata; il tutto è ora magnificamente esposto in una bara di cristallo, nella quale attende la principessa azzurra presso l'University College di Londra. La testa che è oggi possibile vedere, purtroppo, non è la sua: quella vera, rinsecchita come una prugna secca, era oggetto di continui scherzi da parte degli studenti, così ora è conservata altrove sotto vetro e solo raramente mostrata. Nel 2013 il consiglio dell'University College doveva votare su questioni di particolare importanza, così importanti che lo stesso Jeremy Bentham è stato portato alla riunione; non volendo fare favoritismi, si è sistematicamente astenuto.

Il grande obiettivo di Bentham, però, non era finire impagliato né costruire una prigione: aveva un progetto di riforma sociale integrale. Tutta la sua filosofia politica poggia su un unico *principio di felicità* da applicare a tutti i campi (Bentham 2004, 2020):

PF: il bene sta nel causare la maggior felicità al maggior numero di persone possibile

Parla di *felicità*, ma intende *piacere*, il piacere è il modo in cui si distingue il giusto dallo sbagliato: questo è *utilitarismo edonico* nella sua forma più pura. Con la sua freddezza, Bentham mira tramite PF a ridurre le decisioni etiche ad un calcolo matematico: ogni volta che ci si trova in dubbio basta usare il *felicific calculus* per calcolare in modo meccanico quanto piacere o dolore una certa azione causerà ad una certa popolazione, così è possibile decidere automaticamente qual è la migliore tra le opzioni disponibili.

Il bene non può prevedibilmente diminuire la felicità né media né totale: i mostri dell'utilitarismo

Questo PF suona sospetto. Sono in banca, entra un malvivente con la pistola. Cosa devo fare? Non posso mettermi lì seduto a calcolare gli effetti di ogni azione possibile, devo agire in pochi secondi, e dovrò farlo avendo informazioni molto parziali sul contesto e sui possibili effetti futuri della mia azione.

Sembra allora che PF sia applicabile solo in alcuni casi, quelli in cui ho abbastanza tempo da calcolare le conseguenze. Ma: è mai davvero possibile farlo? Se il bene sta nelle conseguenze, devo calcolarle per conoscere il bene di un'azione; ma le conseguenze di un'azione sono infinite e infinitamente complesse, un battito d'ali di una farfalla può causare un tornado dall'altra parte del mondo, le mie facoltà sono troppo limitate per calcolare anche solo un infinitesimo di quello che dovrei calcolare. Se il bene sta nelle conseguenze, io non potrò mai sapere quale azione è bene, non importa il tempo a disposizione.

Al che posso rispondere: è bene quello che *sarebbe* il risultato del calcolo a prescindere dalla mia possibilità di portarlo a termine. Se una persona calcola che uccidere è bene, e uccide, forse non è *responsabile* tanto quanto uno che uccide con la consapevolezza di fare il male, ma la sua azione rimane male: la strada per l'inferno è lastricata di buone intenzioni, una cosa può essere bene o male anche se io non posso sapere che è tale (Sidgwick 1981, Mill 1991). Questa idea è necessaria al *realismo morale*, è l'indipendenza di verità e conoscibilità di cui ho già parlato.

Epperò c'è un gran problema in tale risposta. Il consequenzialismo implica non solo una difficoltà o fallibilità nella conoscenza del bene, implica l'impossibilità di conoscerlo, implica *scetticismo morale*; ma questa è una teoria normativa, dovrebbe permettermi di compiere il bene, non solo definirlo come un ideale irraggiungibile, deve essere un principio pratico e non solo teorico (cfr. Anscombe 1958). Per questo lo scetticismo è indesiderabile in etica tanto quanto lo è in epistemologia: se la giustificazione è *truth-conducive*, una teoria etica deve essere *good-conducive*, deve dirmi quando sono giustificato a compiere un'azione.

Ci sono altri due grandi problemi per l'utilitarismo, associati a quelli che sono con pochi dubbi i tre più grandi eticisti analitici: Derek Parfit, Robert Nozick, e John Rawls.

Parfit, che come filosofo aveva lo stessa freddezza vagamente autistica di Bentham, immagina un mondo nel quale, a seconda di come si agisce, si possano creare quattro situazioni future uguali in ogni aspetto tranne che per la felicità:

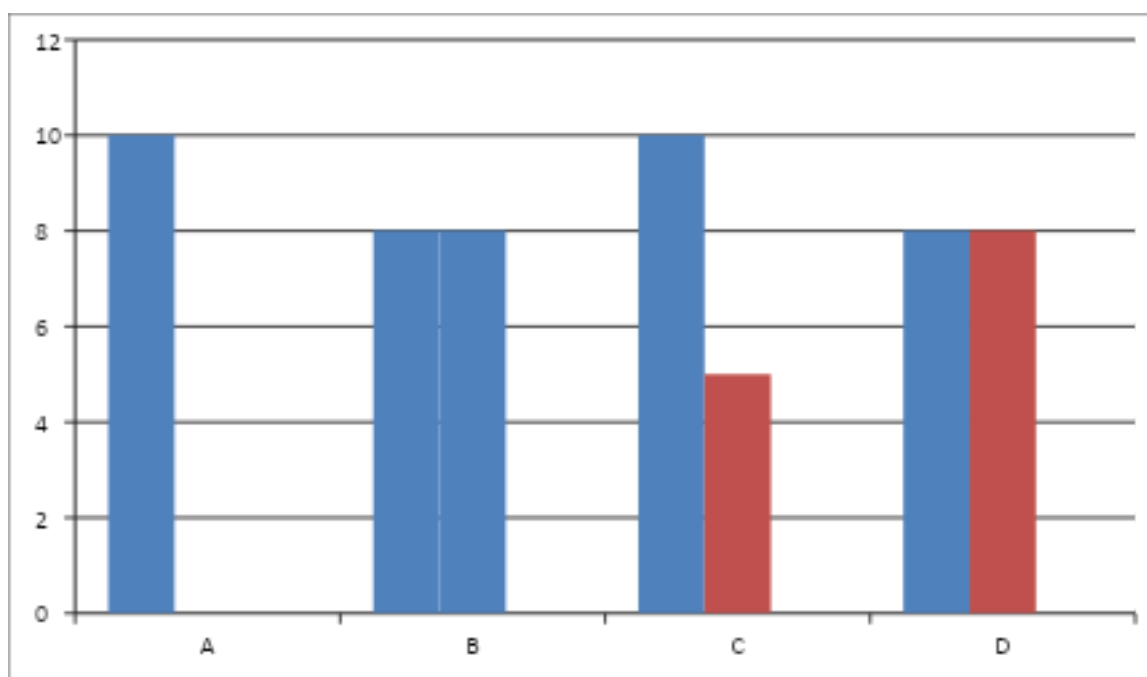


Figura 8

- nel caso A ci sono 5 persone, tutte con un livello di felicità pari a 10 (felicità totale: 50; felicità media: 10);

- nel caso B ci sono 10 persone, tutte con un livello di felicità pari a 8 (felicità totale: 80; felicità media: 8);
- nel caso C ci sono due sottogruppi di 5 persone, uno con felicità 10 e uno con felicità 5 (felicità totale: 75; felicità media: 7,5);
- nel caso D ci sono due sottogruppi di 5 persone, tutti e due con felicità 8 (felicità totale: 80; felicità media: 8).

Verso quale futuro dovrei scegliere di puntare? Noto che:

$$\mathbf{P1} \quad C \geq A$$

La scelta C è buona almeno tanto quanto la A in quanto ci sono 5 persone in più, meno felici delle altre, ma comunque abbastanza felici; quindi, nel rispetto del principio benthamiano, è accettabile una riduzione della felicità media per aumentare la felicità totale.

$$\mathbf{P2} \quad B = D$$

La scelta B è buona tanto quanto la D in quanto hanno uguale felicità media e totale.

$$\mathbf{P3} \quad D > C:$$

la scelta D è migliore della C: è superiore sia per felicità totale che media; quindi

$$\mathbf{C} \quad B > C \geq A:$$

B è migliore di C, e quindi deve essere anche migliore di A.

Per Parfit questo è inaccettabile, secondo lui A deve essere migliore di B. Infatti, se accetto di ridurre la felicità media per aumentare la totale, posso ripetere l'operazione più e più volte fino a creare una popolazione enorme di individui profondamente infelici, un mondo con cento persone tutte con un grado di felicità pari a uno. Questa è la *conclusione ripugnante*, l'idea che il mondo ideale sia un mondo sovrappopolato dove ognuno ha accesso a risorse a malapena sufficienti a rendere la vita appena un poco migliore della morte (Parfit 1989).

Volendo posso alzare l'asticella, posso porre come livello minimo accettabile di felicità non «appena meglio della morte», ma uno stato di benessere più elevato; è però una soluzione arbitraria, e anche ingiusta visto che impedirebbe ad arbitrio la nascita di vite che comunque meritano di essere vissute. Posso allora ipotizzare che la felicità delle persone attualmente esistenti abbia priorità su quella di ipotetici stati futuri, di modo che non sia più legittimo aumentare a dismisura la popolazione per raggiungere uno stato futuro di più alta felicità (Roberts 1998, Ryberg & Tännsjö 2004, Sikora & Barry 1978); questo però non spiega perché, se quello stato fosse presente (foss'anche su un altro pianeta), non sarebbe da considerare come migliore, né spiega perché dovrebbe essere inaccettabile inquinare oggi per rendere il pianeta invivibile domani.

L'unica soluzione, allora, è rifiutare la prima premessa, non posso diminuire la felicità media per aumentare la totale. Forse allora devo fare il contrario: posso accettare di diminuire la felicità totale per aumentare la media, posso passare da un *utilitarismo totale* ad un *utilitarismo medio*. E invece no: Robert Nozick mostra che questa via non è meno paradossale (Nozick 2005).

Nozick immagina di dover scegliere come distribuire le risorse di un certo mondo tra la popolazione. In questo mondo esiste un'entità, un mostro goloso di biscotti, che, qualsiasi cosa esperisca, ne ricava un piacere incomparabile a quello di qualsiasi umano. In tale contesto la presenza di umani oltre al mostro potrebbe aumentare la felicità totale, ma ridurrebbe la media, e per quante risorse si possano dare ad un essere umano comunque la sua felicità non aumenterà mai tanto quanto quella del mostro quando riceve anche solo un biscotto: ne consegue che la scelta migliore è sterminare gli esseri umani per lasciare solo il mostro e così massimizzare la felicità media. Coloro che sono tristi non hanno diritto alla vita, il mondo ideale è un mondo aristocratico nel quale poche persone accentrano tutte le risorse. Un'altra conclusione ripugnante.

John Rawls è stato il più grande eticista analitico, al suo apice era così influente che, nelle riviste accademiche, il suo nome appariva in media una volta ogni tre pagine, a prescindere dall'argomento. Pur non sostenendo l'utilitarismo, offre un modo per gestire il caso del mostro: bisogna aggiungere una clausola, detta *maximin*, che dà maggior valore agli individui più svantaggiati. Se io non ho alcun mezzo per muovermi, la mia prima bicicletta mi sarà molto utile, ma la decima uguale alle altre mi sarà superflua, il valore cala man mano che ricevo biciclette. Similmente, anche se il mostro riceve un piacere enorme da ognuno dei biscotti dai quali è ricoperto, bisogna moltiplicare il piacere datogli dal centesimo biscotto per un indice tale da renderlo inferiore al piacere che una persona senza mangiare riceve dal suo primo biscotto (il che equivale, nel caso di Parfit, a moltiplicare la felicità per un indice che cala all'aumentare della popolazione) (Rawls 2008). Ma questo non è ancora sufficiente, basta ribaltare l'esempio: in tal caso ci sarà un mostro così negletto e disagiato ma con bisogni e desideri così colossali che, per quanti biscotti possa ricevere, la sua condizione non migliora che di un infinitesimo, di modo che dovrei sterminare tutti quelli più felici di lui per assicurargli accesso a tutte le risorse.

Dietro tali esempi ci sono anche interessi politici. L'esempio di Parfit dimostra l'inadeguatezza di una politica socialista e distributista in cui si valuta solo l'uguaglianza economica; l'esempio di Nozick invece fa lo stesso per una politica capitalista in cui si valuta solo la libertà d'impresa (che lui sosteneva fortemente). Ma allora: qual è l'alternativa? Bisogna rivedere PF in modo tale che possa superare entrambe le obiezioni.

La tesi consequenzialista 2: perfezionando l'utilitarismo

Il segretario di Bentham era James Mill. Gli nacque un figlio, John Stuart; d'accordo con Bentham, decise di sottoporlo ad una pedagogia sperimentale estrema e inquietante tanto quanto il progetto del *panopticon*. L'obiettivo dichiarato di James Mill era produrre un genio, non tirar su un figlio ma programmare una macchina perfetta che sarebbe vissuta con il solo scopo di portare a compimento il sistema dell'utilitarismo. E l'esperimento è riuscito bene. Troppo bene. Bentham era stato lui stesso un bambino prodigio (a sette anni intratteneva gli ospiti suonando Handel al violino) ma è nulla rispetto al talento di J. S. Mill: a tre anni leggeva classici latini e greci in lingua originale, a otto anni studiava già la matematica dell'università e aveva scritto un seguito dell'Iliade, a tredici anni Ricardo, il maggior economista dell'epoca, lo invitava in casa propria per confrontarsi su temi di economia politica (Mill 1976). Secondo gli psicologi Maud Merrill e Lewis Terman, J. S. Mill è stato il genio più precoce mai nato (Cox 1926).

Il rapporto tra Mill e Bentham non è limpido. Quando, in un carteggio mai pubblicato, Bentham scrisse a Kant «non sarai tu a fermarmi, John Stuart diventerà più colto sia di me che di te», questo gli rispose «la fiducia nel tuo nuovo apprendista mal posta forse è, così come la tua fiducia nel lato

utilitarista dell'etica». Mill aderiva *in toto* all'utilitarismo, ma non vedeva sempre Bentham di buon occhio, accettava in gran parte il suo sistema ma credeva anche che il principio di felicità così com'era stato espresso avesse fatto più male che bene. Sconvolto da queste critiche, una volta, durante un duello filosofico sul pianeta Mustafar, Bentham gli ha urlato contro «tu eri il prescelto, John! Dovevi portare equilibrio nell'utilitarismo, non lasciarlo nelle tenebre!»

Bentham faceva della felicità un minestrone: se giocare a briscola dà più piacere della poesia, allora giocare a briscola è un bene più grande del leggere o comporre poesia (Bentham 1830). Credere Mill così ingenuo sarebbe mancare di rispetto al suo genio: Mill (1991) mette in gerarchia vari tipi di piacere, dà ai piaceri intellettuali un ruolo più eminente di quello dei piaceri fisici, «è meglio essere un uomo triste che un maiale felice». Il piacere, allora, non è più un criterio autonomo e assoluto del bene e del male, serve anche un *parere esperto* in merito al piacere: la poesia è meglio del giocare a briscola perché tutti coloro che conoscono *entrambi* e *in modo adeguato* i piaceri riconoscono la poesia come superiore. Trattando PF in questo modo Mill riesce a rendere merito di una pluralità di valori, per esempio rende merito del fatto che conoscere la verità è un bene a prescindere dal fatto che a me possa non dare piacere.

Questo compiuto da Mill è il primo passo per salvare PF dalle critiche. E ci sono altri due passi che Mill ha accennato, ma non ha avuto il coraggio di portare fino in fondo.

Il secondo passo sta nel riconoscere che l'uomo, come essere limitato, non può calcolare in ogni circostanza le conseguenze delle sue azioni. Mill si pone il problema ma continua a pensare che il bene di un'azione dipenda dalle sue *effettive* conseguenze anche quando queste non possono essere calcolate; la sua difesa di tale posizione sta nell'idea che, anche se l'uomo è un essere limitato, la società nel suo complesso, lungo tutta la sua Storia, ha avuto tempo sufficiente per maturare un'esperienza tale da sapere cosa ha buone conseguenze e cosa no. Ma si tratta di una difesa fiacca, la società cambia idea ogni anno su cosa sia bene, non sembra avere le idee molto chiare! Per una difesa più energica avrebbe dovuto sostenere che il bene e il male dipendono dalle conseguenze *attese* o almeno, per evitare il rischio di negligenza, dalle conseguenze *prevedibili* in linea di principio.

Il terzo passo vien da sé: non posso calcolare di volta in volta le conseguenze del mio agire, quindi dovrebbe essere per me legittimo scegliere cosa fare basandomi su delle *regole* pratiche che posso applicare senza calcolo (cfr. Hooker 2000); perché in fondo, se la società davvero ha avuto tempo per capire cosa è bene e cosa no, questa comprensione si manifesterà proprio nelle regole di comportamento che impone. L'importante, per rimanere nel solco del consequenzialismo, è che la bontà di tali regole dipenda sempre dalle loro conseguenze (che poi, in certi casi, la regola possa fallire, perché si è calcolato male o perché non può essere così complessa e specifica da essere perfetta in ogni singolo caso, è irrilevante).

Posso riformulare PF così:

PF*: «un'azione è buona se si conforma ad una regola posta con l'intenzione di causare la massima felicità al maggior numero di persone»

o più esplicitamente

PF**: «una regola è buona se è prevedibile che renderà il più felice possibile il maggior numero di persone, un'azione è buona se si conforma a tale regola»

Quello di Bentham era un *utilitarismo dell'atto*, questo è un *utilitarismo della regola e dell'intenzione*, sono passato dal valutare le conseguenze *simpliciter* al valutare le conseguenze che è razionale aspettarsi una certa regola abbia quando *generalmente* rispettata.

Però manca ancora qualche pennellata: ho risolto il problema del calcolo delle conseguenze, ma non ancora quello sollevato dalle risposte di Parfit e Nozick. Le due obiezioni, prese insieme, chiedono di specificare che la felicità media e la felicità totale non possano essere indipendenti, devono aumentare assieme o comunque nessuna delle due deve calare per far aumentare l'altra. Epperò anche specificare ciò non è sufficiente.

Immagino il caso di una persona così depressa da avere una *felicità negativa*, un valore sotto zero così estremo che non raggiungerà mai un valore positivo, e che si convinca che l'unico modo per aumentare la felicità del mondo è suicidarsi. Ha ragione: la sua scomparsa aumenterà più di qualsiasi alternativa sia la felicità media sia la totale, quindi per PF è la scelta più etica da compiere (cfr. Cosculluela 1995). Eppure è una soluzione difficile da accettare. Oppure, immagino un tossicodipendente che, ad ogni nuova dose della sostanza, riceve un piacere sempre più grande, sempre di più, fino a diventare immensamente più grande a tutti gli altri piaceri possibili: per PF la scelta più etica è quella di continuare a farsi nuove dosi perché qualsiasi altra cosa darebbe un piacere più piccolo; ma anche questo è difficile da accettare.

Il problema allora non sta nella distinzione tra felicità media e totale, sta nel criterio della *massimizzazione*, l'idea che la felicità debba essere aumentata *il più possibile*. Ma se si elimina tale criterio, e si accetta un qualsiasi aumento di felicità, anche di piccolo grado, allora il depresso non sarà più spinto a suicidarsi, e il tossicodipendente potrà accontentarsi di altri piaceri (cfr. Slote 1984, Sen 1985, Nussbaum 2000, Elliot & Jamieson 2009).

Ecco il nuovo principio, che chiamerò solo *principio utilitarista*:

PU: un'azione è buona se si conforma ad una regola la cui intenzione è aumentare di un qualsiasi grado la felicità».

La tesi consequenzialista 3: il rule-following paradox richiede esternalismo etico

Il *rule-following paradox* del Kripkenstein è intuitivamente rilevante non solo per l'utilitarismo della regola, ma anche per tutte le altre teorie. L'idea stessa di un'etica presuppone la regolarità delle azioni e quindi la capacità di prevedere quali azioni future saranno legittime e quali no. Il paradosso del Kripkenstein va a danneggiare questo punto mettendo in dubbio la possibilità di poter pervenire al bene tramite regole.

Il paradosso, come la falce della luna, è cornuto. Il primo corno è questo: posta una regola quale «non rubare» è possibile interpretarla legittimamente come *non rubare**, un'interpretazione tale che mi permetta di prendere cose senza pagare; non c'è quindi alcun fatto in merito a quali azioni siano permesse o vietate dalla regola, né posso dire se una mia azione la rispetta o la infrange. Questa parte dal paradosso è risolta nel modo scettico-pragmatico già visto; inoltre, se immagino la regola non come qualcosa da interpretare, ma come una regola che io mi dò e alla cui proposizione ho un accesso diretto, non posso avere dubbi sulla bontà o malvagità delle mie proprie azioni.

Il secondo corno, però, non è così semplice. Parto dall'azione: vedo un uomo prendere una mela, la mette in tasca, e se ne va senza pagare. Posso immaginare anche di conoscere una arbitraria serie di sue azioni passate. La domanda è: a quale regola si sta conformando l'agire di quell'uomo?

Quando uno dice che « $2+2=4$ » quale regola sta esprimendo con il segno «+»? Forse sta esprimendo la regola « $x+y$ è uguale alla somma di x e y », ma anche « $x+y$ è uguale, se $x=y$, alla somma di x e y e, se $x \neq y$, alla loro sottrazione»: non c'è alcun fatto in merito. Quando dice «gavagai» potrebbe intendere *coniglio* come *parti di coniglio*, quando dice «gatto» potrebbe intendere *gatto* come *ciliegia*. Similmente, quando vedo una persona rubare, posso immaginare il suo comportamento come frutto di una regola dannosa quale «ruba più che puoi», o di una più positiva come «ruba ai ricchi per dare ai poveri», o anche di una qualche altra regola tanto complessa quanto grandemente positiva nelle sue conseguenze. Come faccio allora a dire se tale azione è buona o cattiva?

Ogni comportamento può essere accordato con un'infinita serie di regole; per di più, una singola regola può essere interpretata sia come avente buone conseguenze, sia come avente cattive conseguenze; quindi per qualsiasi azione *altrui* esistono sia interpretazioni del mondo tali che l'azione è male, sia interpretazioni del mondo tali che l'azione è bene. Questo, di per sé, non è distruttivo, è accettabile proprio come è accettabile che idee opposte possano essere ugualmente giustificate in diversi sistemi di credenze; bisogna però impedire che ciò sfoci in scetticismo (cfr. Bridges 2014). Com'è possibile evitarlo?

Posso immaginare un universo estremamente semplice, con solo 99 *stati di cose* possibili, 49 felici e 50 dolorosi; immagino quindi di dividere gli stati dolorosi in due gruppi, A e B, e formulo due regole, una tale che vi si conformi qualsiasi azione realizzi uno degli stati felici o uno degli stati negativi del gruppo A, e l'altra uguale ma con gli stati negativi di B. Entrambe queste regole permettono più conseguenze felici che dolorose, quindi sono regole buone; ma sono anche tali che qualsiasi azione si conforma o ad una o all'altra: quindi qualsiasi azione è buona, e ciò è assurdo. Posso obiettare che nessuno può credere entrambe le regole contemporaneamente (perché ne risulterebbe un sistema con più stati dolorosi che felici, e quindi negativo), perciò qualche azione negativa dev'esservi, e dipende da qual è la regola *creduta* nel momento dell'azione. Ma: come faccio io a dire qual è la regola creduta? E perché un'azione fatta senza pensare ad una certa regola non dovrebbe poter essere positiva?

Io credo che la soluzione sia questa: la bontà dipende dalle conseguenze, ma le conseguenze di un'azione non dipendono solo da quell'azione, bensì da come tutti gli uomini agiscono in un certo momento; cioè, dato che le conseguenze delle varie azioni si combinano per portare verso un unico futuro, il bene, non dipende dall'agire dell'individuo, ma da come questo si combina nell'agire complessivo della società. Se io ho delle aspettative sulle conseguenze delle regole senza considerare ciò, allora le mie aspettative sono illegittime. Se invece considero ciò, allora devo riconoscere che la regola che devo seguire non può essere *privata*, non può essere creata da me e valere solo per me, deve invece essere *pubblica*, essere nota e valida per tutti, perché solo così può coordinare l'azione della società, e solo così si possono avere delle aspettative sensate sulle sue conseguenze (cfr. Wittgenstein 2014).

Questo costituisce un'interpretazione esternalista dell'etica. Se la regola etica è interna a colui che agisce, allora non potrò mai sapere se l'azione è prodotta da una buona o da una cattiva regola; ma se la regola è esterna come una legge, allora posso sperare di evitare lo scetticismo sul giudizio. Ciò che rende buona questa regola esterna non è ciò che io immagino come sue conseguenze, ma ciò che la società che pone quella regola si aspetta: l'aspettativa di PU non è personale ma collettiva.

Quando devo giudicare un'azione altrui, posso interpretare la regola come faccio con qualsiasi testo e dedurre quali azioni sono permesse da tale regola e quali no. In questo modo non mi devo chiedere da quale regola può discendere una certa azione, mi devo chiedere solo se la regola pubblica permette oppure no l'azione esaminata; con pragmatismo, presupporrò che la mia interpretazione della regola sia corretta finché non ho motivo di dubitarne, e dirò che gli altri stanno infrangendo la regola solo quando compiono azioni chiaramente vietate.

Il bene non è riducibile alla felicità 1: Matrix

Inizio dai concetti di *felicità totale* e *felicità media*: sono sensati? I singoli uomini possono essere felici o tristi, ma una società, di per sé, non ha una mente, non ha sentimenti; dire «per la società, la morte di cinque persone è cinque volte peggiore della morte di una sola» è insensato ed estraneo al modo in cui normalmente si concepiscono il bene e il male (Scruton 2018).

Nozick continua su questa scia mostrando l'irriducibilità del *bene* a concetti quali *felicità* o *benessere*: se ciò è corretto, allora ogni utilitarismo derivato da PF è insufficiente. Nozick (2005) immagina un mondo *a la Matrix* dove esiste una macchina in grado di simulare ogni esperienza desiderabile: ci si distende sul lettino, la macchina si chiude, un ago entra in modo indolore attraverso la nuca e inizia a stimolare il cervello in modo controllato; chi è dentro la macchina ha l'impressione di vivere una vita da sogno, di correre sui prati, mangiare i cibi più succulenti, vedere meraviglie, la sua famiglia è quella del Mulino Bianco, la mattina si alza già pettinato e il letto si fa su da solo... però nulla di tutto ciò è reale. La domanda è: sarebbe etico collegarsi a questa macchina?

Posso immaginare casi simili più realistici: ci sono davvero droghe che causano una euforia più potente dell'orgasmo, ci sono psicofarmaci che permettono di controllare l'umore, se tutti prendessero regolarmente inibitori del *reuptake* della serotonina forse il mondo sarebbe più rilassato e felice, la gente dovrebbe prendere *Valium* a scopo ricreativo... è etico questo? Se ciò causa felicità e benessere, perché non dovrebbe essere questa la vita ideale?

Sembra proprio che la felicità prodotta da una scelta non sia sufficiente a giustificarla come etica, il *bene* deve prevedere altri fattori: non basta avere l'impressione di vivere una vita felice, bisogna viverla realmente; non è sufficiente causare una reazione chimica nel proprio cervello ingurgitando qualcosa, bisogna essere responsabili in prima persona della propria felicità. Svegliarsi e scoprire che tutto lo studio, il lavoro, i rapporti con le persone amate erano illusioni può essere doloroso, tanto che la scelta istintiva sarebbe quella di ricollegarsi alla macchina, eppure anche così bisogna riconoscere che sarebbe immorale farlo: Neo in Matrix si scollega, e sceglie correttamente. Non tutti i filosofi sono d'accordo (Singer & de Lazari-Radek 2014), non tutti sono pronti per essere scollegati, alcuni sono così assuefatti, così disperatamente dipendenti dalla macchina che lotterebbero per difenderla; ma resta immorale.

Rendere merito dell'inaccettabilità di un omicidio o del collegarsi alla macchina non è complesso, anzi, Mill ha già offerto la soluzione: basta gerarchizzare i piaceri. Mill, però, confronta piacere e piacere, sensazione e sensazione, richiede solo di riconoscere l'intensità dell'esperienza; qui, invece, bisogna dire che un piacere *del quale si ha responsabilità* è superiore ad uno del quale *non la si ha* in modo *indipendente* dalla piacevolezza dell'esperienza, il parere esperto non è sufficiente perché non basta confrontare i piaceri per individuare il più profondo. Si crea così una doppia gerarchizzazione: le azioni sono eticamente gerarchizzate in base alla piacevolezza delle conseguenze, le quali sono a loro volta gerarchizzate secondo un diverso criterio

non-consequenzialista indipendente dalla felicità e dal piacere. L'utilitarismo che ne risulta, allora, è ancora un consequenzialismo?

A questo punto, tanto vale abbandonare il concetto di *felicità* per sostituirlo con quello di *soddisfazione dei bisogni e dei desideri*: in questo modo si rimane nel consequenzialismo, ma non è più legittimo collegarsi alla macchina, essere assetati e sognare di bere non è come bere perché non soddisfa davvero il bisogno. Ecco però che torna Rawls, e mostra che anche questo non è sufficiente.

Rawls (2008) immagina una brillante matematica che, per qualche motivo, conclude che il suo più intimo desiderio è passare tutta la vita a contare i fili d'erba dei prati. Posso dirle che questo desiderio è *momentaneo*, o *mal informato*, o *di secondaria importanza*, e quindi non vale, ma lei fa delle ricerche e si assicura che non è così, quello è davvero il suo più profondo desiderio. Questa scelta di vita la soddisfa, ma sarebbe assurdo considerarla la scelta più etica possibile: è una vita buttata. Posso obiettare: è una cattiva vita perché la matematica così non può soddisfare i suoi bisogni *primari* come mangiare, bere, vestirsi. D'altro canto, un artista negligente verso i suoi bisogni fisici, ma che si dedica fino alla morte a creare un'enorme ed edificante scultura, può essere eticamente ammirabile, addirittura eroico, è virtuoso sacrificare i suoi bisogni per realizzare un desiderio più eminente. Quindi?

Da questi esempi concludo che il bene non può essere qualcosa di banalmente soggettivo, non può essere il fare ciò che uno vuole fare. Molti filosofi, da Moore (1903) a Putnam (1985), hanno sostenuto che il bene non può stare nei desideri perché poi ci si può sempre chiedere «quali desideri è bene avere?», quindi devo comunque porre un criterio non-consequenzialista che gerarchizzi i desideri. Se una persona fa ciò che desidera ma non gerarchizza i suoi desideri allora è egoista, non morale.

Generalizzando, posso arrivare a dire che l'idea che un'azione sia bene o male *solo* in virtù delle conseguenze è controintuitiva. Forse stuprare a morte bambini può aumentare la felicità di una popolazione: PU non obbliga a farlo, però non riesce neanche a spiegare perché tale azione non dovrebbe essere un bene. Il male di un'azione come questa è intrinseco, non può essere ricondotto a soli principi consequenzialisti; e se questo vuol dire che si tratta di un mezzo errato per ottenere quelle conseguenze, che non tutti i mezzi sono ugualmente legittimi, allora devo rifiutare ogni consequenzialismo.

Il bene non è riducibile alla felicità 2: libero arbitrio

Passo infine a ulteriori critiche sulla calcolabilità delle conseguenze, pure di validità generale per l'utilitarismo (cfr. anche Dennett 2015).

Il consequenzialismo si limita a chiedermi di calcolare le conseguenze; l'utilitarismo però, come se questo non fosse già abbastanza difficile, mi chiede di calcolare anche la felicità prodotta sul mondo da queste conseguenze. Io sono incapace di calcolare con sufficiente precisione perfino la mia propria felicità, ogni volta che devo andare a fare la spesa penso che mi annoierò e invece poi sono contento, che io possa stimare addirittura la felicità del mondo è fantascienza, e lo è anche se a calcolare non sono proprio io ma la società nel suo complesso. La fallibilità, di per sé, non è un problema, ma in questo caso è così profonda e pervasiva da rendere impossibile il giustificare le mie credenze sul bene sul male in questo modo.

PU è modesto perché non mi chiede di calcolare con precisione, chiede alla società nel suo complesso solo di calcolare ciò che è umanamente calcolabile. Questo, però, non riesce a chiudere la falla. La teoria deve comunque essere *good-conducive*, calcolo le conseguenze perché voglio andare verso un mondo migliore. Ma il fatto che la società sia in grado di calcolare alcune conseguenze positive di una certa azione non aumenta la probabilità che quell'azione continuerà ad avere conseguenze positive oltre l'orizzonte dei calcoli, e quindi neanche la probabilità di andare verso un mondo migliore. Quindi: se per individuare il bene devo calcolare tutte le conseguenze la teoria risulta in scetticismo, ma se invece mi posso limitare alle conseguenze prevedibili, la teoria è insufficiente, non ho più motivo di credere che io stia andando verso un mondo migliore e quindi non ho neanche più motivo di sostenerla.

Servirebbe qualcosa di diverso, una teoria che, senza calcolare, mi spieghi perché una certa azione risulterà senza dubbio in un effetto desiderabile, o almeno perché non avrà mai effetti catastrofici imprevisi. La fattibilità di una teoria di questo tipo, però, trova un nuovo ostacolo nel libero arbitrio.

Ci sono due possibilità: o c'è libero arbitrio, o non c'è. Immagino non vi sia, cioè che l'universo sia deterministico: posta una certa storia passata, c'è solo un unico futuro possibile. In un caso del genere parlare di etica normativa come di una teoria utile a scegliere come agire non ha senso, non potrò mai scegliere alcunché, agirò sempre e solo così come sono determinato ad agire.

Ora immagino, per assurdo, che anche in un universo determinato vi sia davvero la possibilità di scegliere tra più azioni: epperò, essendo determinato, ad ogni azione dovrebbe seguire un identico futuro, e quindi dal punto di vista delle conseguenze ogni azione sarebbe equivalente alle altre. Se l'universo è così, e io lo so, allora è per me illegittimo avere aspettative sulle conseguenze delle azioni tali da poter valutarne una in modo diverso dalle altre: quindi l'utilitarismo non mi permette di fare scelte. Se invece l'universo è così, ma non lo credo, l'utilitarismo porta a fare valutazioni false.

Passo allora all'altra possibilità: c'è libero arbitrio. Se è così, deve essere genuinamente possibile scegliere tra azioni alternative, il che vuol dire che l'universo ha più futuri possibili. Se però sono libero in modo tale che il comportamento non risulti né determinato, né casuale, né riducibile a leggi statistiche, allora non ho alcun motivo per credere che uno qualsiasi di questi futuri possibili sia più probabile dell'altro. Ipotizzo dunque che esistano più futuri possibili, ma solo nel senso che l'universo evolve in modo casuale. Similmente, non c'è un futuro più probabile di un altro, e non è quindi neanche possibile favorire un'azione sulle altre poiché ognuna avrà sia futuri positivi che futuri negativi, ogni diversa aspettativa sulle conseguenze di un'azione sarebbe per me illegittima.

Solo in un caso l'utilitarismo funziona: il caso in cui l'uomo è davvero libero, ma senza che ciò impedisca di individuare dei futuri possibili più probabili degli altri. L'utilitarista deve sperare che si dia questo caso, e non pare improbabile (se l'uomo è libero, comunque è razionale e quindi parzialmente prevedibile, ogni bambino sa come reagiranno i genitori ai suoi malanni). Ma questo lo approfondirò nella sezione di metafisica.

La tesi deontologica: regole, Legge naturale, religione

Mill fu il padrino di Russell, ed entrambi consideravano l'utilitarismo come la cifra etica della loro tradizione filosofica, contro la deontologia dei continentali (Russell 2004). Oggi le cose sono

cambiate: molti dei più grandi eticisti analitici quali Nagel, Rawls, il conservatore Scruton, hanno rifiutato l'utilitarismo proprio in favore della deontologia.

La deontologia pone il valore nell'azione a prescindere dalle conseguenze. Questo permette di distinguere azioni che sono male anche quando le loro conseguenze sono molto positive, oppure azioni che sono bene anche se hanno conseguenze negative. Ciò è necessario per rendere merito del valore della *giustizia*: forse torturare un innocente renderà gaudiosa tutta la popolazione, ma questo è *ingiusto* perché l'innocente *non merita* di subire ciò e quindi non va fatto; forse mandare in carcere l'assassino diminuirà la sua felicità e non aumenterà quella di alcun altro, ma *se lo merita* ed è *giusto* che sia punito.

Nella deontologia, il valore dell'azione sta nel suo conformarsi a certe regole. Se in PU il valore delle regole sta nelle loro conseguenze, qui dipende tutto dal senso del dovere. Posso dire così: «è una regola quella che ho il dovere di rispettare»; se non ho dovere di rispettarla non è una regola in primo luogo.

Dunque: quali sono le regole? Il concetto di dovere si collega subito al concetto di diritto e di Legge: una legge è tale se chi la emette ha il potere di porre sulle mie spalle il dovere di rispettarla. E questo è un primo tentativo di risposta: le regole sono quelle determinate dalla Legge; e quindi: «un'azione è buona se rispetta della Legge». Una morale per avvocati.

È ovvio quanto questa definizione sia semplicistica. In alcuni Stati c'è ancora la pena di morte, e dico "va abolita perché è male", non dico certo "diventerà male quando sarà abolita!" In alcuni Paesi è tradizione sposare le bambine (in Iran la maggiore età è a 9 anni), ma non dico "li è bene e qui è male", il bene non conosce confini. Queste obiezioni sono così facili perché, come faceva Bentham, presuppongono che la Legge sia una convenzione: e poiché non si accetta che anche il bene e il male siano semplici costruzioni (chi direbbe che stuprare bambini innocenti è male solo per convenzione?), è ovvio che vi sarà sempre uno scarto tra Legge ed etica, sarà sempre possibile immaginare una legge che obbliga a fare qualcosa di palesemente immorale, oppure un atto immorale che però non può essere vietato per Legge (la bugia di un bambino è immorale, ma non lo si può multare per questo). Ma questo presupposto è discutibile: un filosofo non è obbligato a considerare Legge «quella in vigore in un certo Stato in un certo momento», può averne una concezione più astratta e non-convenzionale.

Ugo Grozio è stato un genio la cui precocità è paragonabile a quella di Mill (Cox 1926): entrò all'università a undici anni, a quindi anni iniziò a compiere missioni diplomatiche e fu presentato alla corte di Francia come un miracolo, l'anno dopo divenne avvocato e a diciotto gli Stati d'Olanda gli commissionarono la stesura della loro Storia ufficiale. Interessandosi di diritto internazionale e bellico, si trovò nella necessità di far poggiare la sua teoria legale su delle fondamenta che fossero indipendenti dalle credenze, dai sentimenti, dalle tradizioni dei singoli popoli coinvolti, di modo che il diritto derivatone fosse accettabile da tutti. L'idea, già presente nella filosofia romana (Cicerone 2019) ma particolarmente sviluppata da Grozio (2002), è che esista un tipo di Legge indipendente dalle specifiche leggi dei singoli Stati, una Legge che ha le sue radici nella natura del mondo e dell'uomo e che si riflette nel fatto che tutti i sistemi legali, in tutti i luoghi e in tutti i tempi, hanno riconosciuto una serie di norme (non uccidere, non rubare...) come *universali, innegabili, necessarie* per ragione: tale è la *Legge naturale*.

Ci sono più basi possibili per la Legge naturale. Un esempio è la teoria dell'evoluzione: la selezione naturale ha favorito certi valori, l'altruismo è stato vantaggioso per la sopravvivenza, i geni associati a tale comportamento sono stati tramandati e ora questo è un valore basico, sentito da tutti perché

dipende dal modo in cui gli uomini sono fatti e non da ciò che possono sapere o credere. Oppure, anche senza appellarsi all'evoluzione si può individuare una *natura umana*: il furto è male perché va contro legami sociali fondamentali, questo male sta non nelle conseguenze della rottura dei legami ma nel fatto che l'uomo ha una natura sociale e il furto va contro tale natura. Un'altra possibilità ancora, almeno per chi ha fede nel progresso, è che prove ed errori nei tentativi di risoluzione dei conflitti sociali e internazionali portino a perfezionare le norme esistenti fino alla scoperta delle norme *così come sarebbero dovute essere*.

L'epitome della Legge naturale, però, è quella riflessa nell'*etica religiosa*. Grozio dava un fondamento teologico alla Legge ma pensava, come già il francescano Duns Scoto (2005), che la Legge naturale sarebbe rimasta valida anche ipotizzando l'inesistenza di Dio. Anscombe non è d'accordo, secondo lei né il consequenzialismo né la deontologia possono essere sensati se si prescinde dal concetto della volontà di Dio. Anscombe (1958) fa un'osservazione molto acuta: nel corso della Storia, i concetti etici si sono sviluppati di pari passo con le teorie teologiche; al contrario, l'enorme sviluppo nella comprensione scientifica della Natura degli ultimi secoli non ha portato alcun progresso nell'etica. Se davvero il bene stesse nell'evoluzione, nella natura umana, nella Storia del diritto, questo sarebbe inspiegabile; al contrario, si può spiegare ipotizzando che i concetti etici siano concetti teologici sotto mentite spoglie.

Infatti ho proposto più fondamenti per la Legge, ma nessuno di questi, tranne Dio, sembra poter giustificare l'idea che la Legge sia normativa (Adams 1987). Cosa giustifica il mio ritenere l'altruismo un comportamento etico? Se rispondo l'evoluzione, o il fatto che il mio corpo è strutturato in un certo modo, poi devo chiedermi: e perché ciò lo renderebbe etico? La Natura non può obbligarmi ad essere altruista perché è neutra, non è in sé morale. Ma se la Legge naturale è decretata da un legislatore già morale nella sua essenza, imparziale e indiscutibile, allora sì che capisco perché devo essere altruista. E allora capisco anche perché la morale si può fondare solo sul concetto di Dio.

Contro tale teoria si solleva un'obiezione risalente, come sempre, a Platone: il *dilemma di Eutifrone* (in Platone 2009) Il bene sta nella volontà di Dio? È un problema: se Dio vuole certe cose perché queste sono bene, allora il bene è indipendente da Dio e la teoria è falsa; ma se una cosa è bene perché voluta da Dio, allora il bene è arbitrario, non c'è alcun motivo per il quale Dio non avrebbe potuto volere tutt'altro, e quindi neanche alcun motivo per il quale quelle azioni siano davvero bene. A tale dilemma si può dare una risposta teologica (Harriet 2011). Per tradizione Dio è inteso come massimamente *semplice*, senza alcuna parte; se è così semplice, allora non posso distinguere in Lui *sostanze* e *attributi*, cioè non posso distinguere *ciò che esso è* dal *modo in cui è*: posso distinguere il fiore del suo colore, ma non posso distinguere Dio dalla sua bontà. Per questo motivo è scorretto immaginare una Legge naturale così fondata che non sia buona; se la Legge non è buona, non è Legge, è insensato chiedersi se la Legge sia buona o cattiva.

L'etica religiosa intesa come forma di deontologia basata sulla Legge naturale è indipendente dalle etiche delle specifiche religioni, le quali sono difficili da ricondurre ad un'unica teoria filosofica: per esempio, l'etica cattolica è consequenzialista (punta a portare il *Regno di Dio* in terra, a massimizzare l'amore) (Fletcher 1997), è deontologica (ha i *dieci comandamenti*), ed è un'etica della virtù (prevede virtù *cardinali* e *teologali*). Inoltre, mentre l'etica di una religione presuppone credenza, l'etica religiosa deontologica può essere sostenuta anche da un ateo: uno può sostenere che il concetto di bene è equivalente al concetto dell'obbedienza a Dio anche se non crede che Dio esista.

Il bene è un dovere universalizzabile, sensibile alle motivazioni e sempre rispettabile

La variante di maggior successo della deontologia è l'*etica kantiana*. Quella che espongo ora è una teoria basata su quella di Kant (1785, 1790, 1792), ma che contiene anche molti elementi personali e idee provenienti dagli altri pensatori che hanno continuato, e tutt'oggi continuano, questa tradizione.

Inizio col concetto di *bene*, per controllare come si collega all'imperativo categorico. È intuitivo che vi sia un legame tra il *bene* e il *dovere*: se ho il dovere di fare qualcosa non posso considerare questo un male, e se so che una cosa è bene sento anche il dovere di farlo.

Nell'etica kantiana il bene ha due caratteristiche principali. La prima è l'*intrinsecità*; quindi non può essere un dovere estrinseco, relativo a condizioni esterne. La seconda caratteristica del bene sta nel fatto che «la sua aggiunta non può rendere eticamente peggiore una situazione»; cioè, non si può usare il bene *per il male*. Questo permette di escludere varie presunte definizioni di «bene». Immagino una città nella quale vi sia un'ampia fetta di poveri costretta a vivere sulla strada, e ora vi aggiungo un sadico che gode nel vedere la sofferenza di questi senz'altro. La presenza del sadico, per un utilitarista, aumenta il piacere medio e totale, ma è ovvio che la sua perversione peggiora la condizione etica della città: quindi il bene non può stare nel piacere. È come dire che il bene non può stare nella fedeltà, perché la fedeltà ad un uomo malvagio si tradurrebbe in azioni malvagie (questo non implica che l'essere fedeli sia male, ma solo che il bene non è riducibile alla fedeltà).

Questa seconda caratteristica mostra che il bene è *sensibile* alle motivazioni: fare la carità è cosa buona, ma se la faccio solo per apparire munifico agli occhi degli altri, non per aiutare, non è più un bene, il mio fare la carità peggiora la condizione etica del mondo. Altro esempio: il mio arcirivale sta camminando sui binari quando sopraggiunge il treno, io non voglio che muoia (pur odiandolo) e so che potrei salvarlo urlandogli di spostarsi, epperò, essendo mio rivale, voglio fargli del male, quindi decido di salvarlo spingendolo in modo tale da storcergli una caviglia: devo salvarlo, l'ho fatto, e questo è bene, ma dovevo scegliere un mezzo migliore, l'aver una cattiva motivazione mi ha portato a non farlo, e questo è male (cfr. Bronner & Goldstein 2018, Di Nucci 2014). Ancora: ho promesso di aiutare l'amico, viene e mi chiede aiuto, vuole che io torni indietro nel tempo per salvare sua moglie morta d'incidente anni prima: se io non lo aiuto perché ciò è impossibile e glielo spiego, non faccio un male, ma se io non so che è impossibile (anche se lo è) e non provo a prestare aiuto semplicemente perché non mi va, faccio un male, peggioro la situazione etica.

In quanto a motivazioni, per l'etica kantiana, il bene deve essere *libero*: tutto ciò che ha una motivazione estrinseca, tutto ciò che è motivato dalla volontà di ottenere certe conseguenze, o dalle pressioni sociali, o dalla paura delle punizioni, non può essere il bene, il bene deve avere una motivazione intrinseca.

Questa motivazione particolare caratteristica del bene è la *buona volontà*, la volontà di chi agisce per rispettare un dovere morale in modo indipendente da conseguenze e interessi personali, senza alcun egoismo. Questo punto è spesso frainteso: si potrebbe intendere che chi si prende cura di un padre anziano fa il bene solo se lo fa per dovere morale e non per amore (cfr. Baron 2003). Nessun pensatore ha mai voluto dire questo, sono due diversi livelli del discorso: il figlio ama il padre, e proprio tale sentimento gli dà il dovere di prendersi cura della persona amata; quindi, come *motivazione di primo livello*, agisce per amore, ma proprio così facendo implica che, come *motivazione di secondo livello*, agisce per dovere morale. Epperò, affinché l'azione sia etica, il dovere deve essere sufficiente e l'amore non-necessario: l'amore può fare da base al dovere, ma il dovere è

indipendente e può avere più basi; se il figlio odia il padre ma deve prendersene cura perché è medico allora, se è etico, lo farà comunque. Il dovere morale manifesta la buona volontà, e l'amore no, perché se si desse la motivazione di secondo livello ma non quella di primo si dovrebbe comunque fare ciò che si fa (cfr. Robert & Cureton 2021).

Or dunque, questo dovere morale associato alla buona volontà, libero, assoluto, intrinseco, indipendente dagli obiettivi personali, deve essere quello dell'imperativo categorico: il bene ne richiede il rispetto. Ma cos'è esattamente un imperativo categorico?

Kant, analizzando il concetto di *dovere*, distingue l'*imperativo categorico* dai vari *imperativi ipotetici*. Un imperativo ipotetico è una regola che io devo seguire *se* voglio raggiungere un certo scopo: sono sì doveri, ma relativi, dipendenti da desideri e interessi. La forma di un imperativo ipotetico è sempre «se voglio *x* devo fare *y*»: se voglio dissetarmi devo bere, se voglio acculturarmi devo studiare... L'imperativo categorico, invece, è insensibile a desideri e interessi, è valido in tutti i contesti perché esprime un bene indipendente dagli scopi (anche se può fare da scopo). La sua forma è «devo fare *x*», senza *se* e senza *ma*.

Come Peirce con la sua massima, anche Kant definisce l'imperativo categorico più volte, ogni volta da una diversa angolazione:

- dal punto di vista dell'*universalizzabilità*, l'imperativo categorico chiede di «agire in accordo con una legge che può diventare universale»; cioè, l'imperativo categorico è una legge che non porta a contraddizioni neppure se rispettata da tutti gli uomini in tutti i tempi, e posso individuarlo chiedendomi «cosa accadrebbe se facessero tutti così?»;
- dal punto di vista dell'*umanità*, l'imperativo categorico chiede di «trattare gli esseri umani, sia presi singolarmente che in gruppo, sempre come fini», cioè devo rispettare i loro obiettivi, non possono mai trattarli solo come mezzi per realizzare miei o altrui fini; per esempio, non è legittimo usare il lavoro di un'altra persona per arricchirmi, o le sue idee per diventare famoso, o il suo corpo per ripararmi da un proiettile, a meno che non sia quella stessa persona ad acconsentire;
- dal punto di vista del *Regno dei Fini*, l'imperativo categorico chiede di «agire in prima persona come se le leggi che si stanno seguendo fossero leggi di un regno universale»; cioè devo agire secondo una legge che sarebbe accettata da una comunità di individui razionali tale da accettare solo leggi che trattano ogni individuo come un fine; cioè l'imperativo categorico è una legge universale che non infrange la libertà di alcuno in quanto tratta tutti come fini e non solo come mezzi.

Questi tre principi non sono tre tipi d'imperativo, sono tutti elementi dell'imperativo categorico. Com'è possibile che tre principi così diversi siano parti dello stesso concetto? Bisogna considerare che l'universalizzabilità riguarda anzitutto i *doveri* e l'umanità riguarda anzitutto i *diritti*: io ho il dovere di fare ciò che la ragione mostra essere necessario tramite l'universalizzazione, e ognuno ha il diritto di non essere usato come mezzo; ma poiché, più spesso che no, i miei diritti sono i doveri altrui e i diritti altrui sono i miei doveri, universalizzabilità e umanità sono in stretta sintonia. Altre teorie deontologiche tendono a ridurre l'etica ad uno solo dei due elementi, ma l'etica kantiana considera diritti e doveri insieme, anche se può dare loro un diverso peso in diverse circostanze. Il *Regno dei Fini* è una formulazione *contrattualistica* di quanto già espresso dagli altri due principi: le leggi non dipendono dalla decisione di una persona, le leggi sono le regole che un'intera società posta in una condizione ideale (perfettamente razionale, libera, mai manipolata...) accetterebbe; le leggi non valgono per individui, valgono per società nel loro complesso. Questo è in linea con l'*esternalismo etico* usato per evitare lo scetticismo legato al *rule-following paradox*.

C'è un ultimo principio necessario a comprendere l'etica kantiana, quello del *se devo allora posso* ($D \rightarrow P$). Sartre (1946) fa un esempio classico di *dilemma morale*: un ragazzo ha il dovere di andare a combattere al fronte per difendere la patria e vendicare il fratello, però ha anche il dovere di restare a casa ad accudire la madre sola, e non può rispettare uno di questi doveri senza violare anche l'altro. È possibile che siano entrambi imperativi categorici in conflitto? Se si considera il principio detto, no, non è possibile: poiché il ragazzo non può sia andare al fronte sia accudire la madre, non ha neanche un dovere verso entrambe le cose. Nel valutare le varie possibilità d'azione, *se io sapessi* che un certo obiettivo è impossibile, non posso razionalmente sceglierlo, e poiché l'etica kantiana è razionalistica non può imporlo come dovere. Se io prometto di andare a prendere la mia fidanzata alle nove, ma mi sveglio tardi e diventa impossibile arrivare in orario, *avevo* il dovere di andare a prenderla *in orario*, ora non lo ho più (ma cfr. McConnell 1989).

L'imperativo categorico, quindi, è sempre consistente. Su come ciò risolva la questione del ragazzo di Sartre si può discutere, perché non è chiaro quale dei due doveri sia quello categorico, se ve n'è uno: forse l'analisi razionale può portare a gerarchizzare le due possibilità, oppure si può scegliere tramite un altro criterio senza venir meno ad alcun dovere qualsiasi sia la scelta.

Ora manca solo da vedere come fare a capire se una certa azione è richiesta dall'imperativo categorico, e quindi è bene, oppure no.

Il metodo ideale per derivare una legge etica dall'imperativo categorico è dimostrare che l'idea che il comportamento preso in esame sia universale è logicamente incoerente, cioè dimostrare che non rispetta il principio di universalizzabilità. Questo esempio viene da Pauline Kleingeld (2017), filosofa dell'università di Groningen. Mi chiedo «è legittimo per me fare una promessa che so non poter mantenere?» Per il principio di universalizzabilità io posso voler fare qualcosa solo se posso volere anche che quel comportamento sia universale, posso voler ingannare solo se posso anche volere che «tutti possano fare promesse e non mantenerle mai». Ma in un mondo in cui tutti credono che sia legittimo fare promesse e non mantenerle nessuno si aspetterebbe che chi promette manterrà la promessa, in un mondo in cui tutti possono ingannare nessuno può ingannare. C'è allora una contraddizione: io voglio ingannare, ma voglio anche che tutti ingannino e quindi che sia impossibile ingannare, il che è assurdo. Quindi l'azione non è universalizzabile, è da vietare: l'imperativo categorico afferma «non mentire».

Un altro aspetto di tale metodo sta nel dimostrare che l'idea che il comportamento preso in esame sia svolto con il consenso delle altre persone coinvolte è logicamente incoerente, cioè dimostrare che viola il principio di umanità. Mi chiedo di nuovo «è legittimo ingannare?» Lo è se posso dire che «inganno con il consenso altrui». Ma io posso ingannare qualcuno solo se questo non sa di essere ingannato, non può darmi il consenso ad ingannare, quindi la frase è insensata e l'inganno è inaccettabile.

Con la prima argomentazione ho dimostrato l'impossibilità di un dovere a ingannare e l'esistenza di un dovere a dire la verità, con questa seconda ho dimostrato l'impossibilità di un diritto a ingannare e l'esistenza di un diritto a non essere ingannato.

L'imperativo categorico così si manifesta tramite *contraddizioni della volontà*, rifiuta tutti i casi in cui volere una cosa e volere che sia universale e fatta col consenso altrui risulta in una contraddizione logica: questi sono *doveri perfetti*. Tali doveri sono però rari, più spesso l'universalizzazione porta ad una *contraddizione pragmatica*, cioè porta ad un contesto estremo razionalmente impossibile da desiderare. L'imperativo categorico rifiuta anche questo tipo di azioni, creando *doveri imperfetti* che non è necessario imporre in modo forte come i perfetti.

Se quelli visti sopra sono doveri perfetti, ecco un esempio di dovere imperfetto. Mi chiedo «è legittimo, per me, non fare figli?» Per saperlo devo provare a universalizzare il comportamento, cioè mi chiedo «cosa accadrebbe se nessun umano facesse più figli?» Desiderare che nessuno faccia figli non è di per sé illogico, ma porta ad un mondo in cui gli esseri umani sono estinti, una conseguenza impossibile da desiderare. Al contrario mi chiedo «è legittimo fare figli? Cosa accadrebbe se tutti ne sfornassero a palate?» e anche in questo caso ottengo una conseguenza indesiderabile, un mondo sovrappopolato e senza risorse nel quale la vita è peggiore della morte. Quindi, se si può, è illegittimo non fare figli, ma è anche illegittimo riprodursi senza criteri e limiti.

Anche questo metodo d'analisi dell'imperativo categorico è stato più volte frainteso. Hegel (1807) obietta che se tutti facessero la carità non ci sarebbero più poveri e fare la carità diventerebbe impossibile, quindi per l'imperativo categorico è male. Ma dire «voglio fare la carità, e voglio che tutti la facciano» non costituisce una contraddizione: infatti io voglio fare la carità proprio perché spero non ci siano più poveri, quindi «se la fanno tutti, nessuno può più farla» è coerente con la mia volontà, è proprio lo stato al quale punto. Al contrario un desiderio come «voglio che tutti ingannino» è incoerente, sarebbe assurdo se io volessi ingannare affinché nessuno inganni.

Alasdair MacIntyre, eticista cristiano ex-marxista, fa l'esempio di una regola universalizzabile ma immorale, «mantieni tutte le tue promesse tranne una» (MacIntyre 2007). Epperò il *Regno dei Fini* afferma che la giustizia non fa eccezioni, ingannare è immorale proprio perché è un qualcosa di attuabile solo se eccezionale, solo se lo fa alcuno in alcuni casi ma non ognuno in tutti i casi. Quindi una legge che prevede al suo interno una eccezione come questa va rifiutata a prescindere.

Questi fraintendimenti sono in parte causati da una non sufficiente analisi del senso delle regole. Penso a «guida a sinistra»: è accettabile che tutti guidino a sinistra, pare universalizzabile, quindi è per me legittimo guidare contromano; e invece no, perché il senso della regola è «voglio guidare in senso contrario rispetto agli altri», e va rifiutata perché prevede un'eccezione e non è universalizzabile.

Mill (1991) nota correttamente che in esempi come questi la bontà o malvagità di un'azione viene giudicata solo in virtù delle conseguenze. Questo non significa che la deontologia si risolva nel consequenzialismo. L'etica kantiana non è insensibile alle conseguenze, ed esiste anche un *sensu etico* che permette di intuire la loro bontà (l'imperativo categorico è infatti ciò che un essere libero e razionale sentirebbe di dover o non dover fare) (cfr. Hume 2009), ma questa bontà non sta nella *sensazione* di piacere o felicità, bensì nell'*incompatibilità con la ragione*: il consequenzialismo è empirico e la deontologia è razionalista. L'etica kantiana, quindi, non richiede di calcolare le effettive conseguenze di un'azione, chiede solo di calcolare l'effetto che avrebbe nel caso più estremo, il caso in cui tutta l'umanità rispetti sempre e comunque la relativa regola di comportamento (non che sia un calcolo facile, comunque).

Il dovere è contestuale: obblighi e divieti

Le nozioni di *premio* e *punizione* sono rilevanti anche per l'etica. C'è un concetto nella filosofia inglese, quello di «moral desert», deserto morale (un gioco di parole che rimanda a *to deserve*), o anche «moral dessert» (perché la torta è un premio), traducibile semplicemente come *merito* (Nozick 2005, Rawls 2008). È vero che la paura della punizione o il desiderio del premio non possono mai *motivare* l'azione etica (cfr. Falk 1947), l'uomo etico agirebbe come agisce comunque; ma il merito richiede che ad un certo dovere corrispondano certe conseguenze. Chi ruba merita di

essere punito, chi è eroico merita di essere celebrato: l'infrangere un dovere etico dà alla società il diritto di punire e a sé il dovere di accettare la punizione, l'essere particolarmente etici dà a sé il diritto ad essere premiati e alla società il dovere di premiare.

Diritti e doveri, implicando premi e punizioni, creano *obblighi* e *divieti*. Questi concetti sono fondamentali sia per la Legge che per la deontologia kantiana; la Legge, però, non collega mai obblighi e divieti ai concetti di *bene* e *male* o *giusto* e *sbagliato*: la Legge si limita a istituire conseguenze (spesso punizioni) al darsi di certi casi. La Legge non dice che rubare è male, dice che chi ruba va in carcere; non dice che pagare le tasse in tempo è giusto, dice che chi ritarda pagherà di più. Quindi, ciò che rende un'azione un obbligo o un divieto è la reazione della società al compiere o non-compiere quell'azione.

Nella sezione precedente ho riflettuto su come individuare diritti e doveri, bene e male. Come individuare obblighi e doveri da un punto di vista morale? Tramite l'imperativo categorico, davanti ad un'azione, mi chiedo «cosa accadrebbe se tutti facessero così?» (*contesto uno*) Devo però anche chiedermi in negativo «cosa accadrebbe se nessuno lo facesse?» (*contesto due*) Ottengo così quattro casi possibili.

Se il contesto uno è inaccettabile, ma il contesto due è accettabile, allora l'azione esaminata è da vietare: un mondo in cui tutti rubano sarebbe contraddittorio (il furto presuppone la proprietà privata, se tutti volessero rubare non vi sarebbe proprietà privata e quindi nessuno potrebbe rubare), ma uno in cui nessuno ruba è coerente e desiderabile, quindi bisogna evitare che il furto dilaghi ma non è problematico che il furto scompaia dal mondo, *ergo* il furto è da vietare.

Se il contesto uno è accettabile, ma il contesto due è inaccettabile, allora l'azione esaminata è da obbligare: un mondo in cui tutti i genitori che possono farlo danno adeguato nutrimento ai figli neonati è coerente e desiderabile, ma un mondo in cui nessun genitore lo fa porterebbe all'estinzione ed è da rifiutare, quindi non è problematico che questo diventi un comportamento universale ma bisogna evitare a tutti i costi che nessuno lo attui più, quindi è da rendere obbligatorio.

Se sia il contesto uno che il due sono accettabili, allora non c'è regolamentazione: un mondo in cui ognuno si rilassa guardando un film la sera è coerente tanto quanto un mondo in cui nessuno lo fa, non c'è interesse a evitare alcuno dei due mondi, quindi scegliere se attuare l'azione oppure no può essere lasciato ai desideri individuali. Azioni di questo tipo non possono costituire doveri, ma possono comunque risultare bene o male secondo altri criteri che è possibile aggiungere al sistema deontologico (l'etica kantiana prevede normalmente anche una serie di virtù, si veda Nussbaum 1999). Queste azioni sono d'interesse anche per il consequenzialismo: infatti sono tali che, anche senza calcolare, si ha motivo di credere (pur in modo fallibile) che non risulteranno mai in conseguenze catastrofiche al di là dell'orizzonte della prevedibilità, rendendo così potenzialmente *good-conducive* il principio utilitarista basato sull'intenzione (pur permanendo gli altri problemi visti).

Se sia il contesto uno che il due sono inaccettabili, allora si sta valutando un comportamento complesso: non può essere obbligato perché porterebbe al contesto uno, né vietato perché porterebbe al due, c'è invece bisogno di un'azione dinamica, è necessario porre dei *livelli di guardia* per mantenere la percentuale di persone che attua quel comportamento all'interno di un certo intervallo. L'esempio del fare figli è di questo tipo: non si può sovrappopolare il mondo facendo figli a nastro, non si possono non fare fino all'estinzione, bisogna mantenere la popolazione ad un livello compatibile con le risorse disponibili attuando di volta in volta politiche adeguate. In questo caso la società ha diritto a sopravvivere, ma nessun individuo ha specifici obblighi o divieti in

merito: man mano che ci si allontana dal livello ottimale ognuno è sempre più pressato ad impegnarsi affinché la società mantenga l'equilibrio.

Se infrango un dovere sono responsabile delle sue conseguenze: la sorte morale

Ho analizzato finora diritti e doveri, bene e male, obblighi e divieti: non ho ancora detto abbastanza su premi e punizioni, devo vedere com'è che la deontologia d'ispirazione kantiana gestisce meriti e colpe e permette di decidere chi premiare o punire e quando.

Primo esempio. Due uomini, chiamati Uno e Due, passano la notte in discoteca, si ubriacano, la mattina si mettono alla guida, entrambi affrontano una curva, sbandano, finiscono fuori strada contro un albero. Uno distrugge la macchina e stop; nel caso di Due, invece, sotto l'albero c'è un bambino che gioca, viene investito e ucciso. Uno non subisce particolari conseguenze a parte una multa, Due viene accusato di omicidio stradale e arrestato, e mi pare giusto così. Perché?

Questo è il problema della *sorte morale*, analizzato da due dei grandi eticisti di questa epoca, Williams (1981) e Nagel (1979).

Ho detto che «se devo, allora posso». Sembra logico, allora, che io non possa essere accusato di colpe che non avrei potuto evitare: non posso essere colpevole della morte di un paziente incurabile. Lo chiamo *principio di controllo*:

PCo: uno può avere meriti o colpe solo per ciò che è *in suo controllo*

Uno e Due hanno agito nello stesso modo, entrambi hanno guidato in stato d'ebbrezza, nessuno dei due voleva uccidere, e che sotto uno dei due alberi ci fosse un bambino è stato frutto del caso, è stata una sfortuna. Se PCo è vero, allora sono entrambi responsabili per l'essersi messi alla guida ubriachi, ma nessuno dei due può essere responsabile della morte del bambino in quanto la sua presenza era imprevedibile e al di fuori del controllo di entrambi. Quindi sono ugualmente colpevoli, il fatto che Due sia un assassino non lo rende più colpevole di Uno. Ma questo è fortemente controintuitivo, è chiaro che Due merita una maggior punizione, e PCo deve essere falso. Ma com'è possibile che fattori indipendenti dall'azione influenzino la responsabilità per quell'azione?

Per il consequenzialismo la risposta è ovvia: PCo è falso perché il bene e il male dipendono dall'entità delle conseguenze, poiché in un caso ci sono state conseguenze più gravi c'è anche maggiore colpa. Nella deontologia la soluzione non è così banale: Uno e Due hanno rispettato o infranto le stesse regole, le conseguenze sono irrilevanti, e devono quindi essere puniti in modo eguale. Questo è assurdo. È possibile rendere merito della differenza senza appellarsi alle conseguenze e senza rinunciare a PCo?

Secondo esempio. Due uomini, chiamati Tre e Quattro, sviluppano l'ossessione di dover uccidere il Presidente. Tre si presenta al corteo, punta la pistola, spara, ma manca il bersaglio e viene subito arrestato; il corteo continua, pochi minuti dopo Quattro esce dalla folla, punta la pistola, spara e uccide il Presidente. Tre viene condannato a dieci anni, Quattro all'ergastolo.

Il *tentato* omicidio, di per sé, non ha conseguenze, cionondimeno è male e non certo pari al dolce far niente: quindi il giudizio etico dipende non solo dalle conseguenze, ma anche dalle intenzioni

(ma cfr. Rachels 1975). L'intenzione di Tre e Quattro, però, è identica. Che il proiettile andasse a segno, e non fosse invece schivato, deviato, o incassato da una guardia del corpo, è qualcosa che sta al di fuori della possibilità di controllo di entrambi. Quindi, che uno spari per uccidere e vada a segno, o spari per uccidere ma fallisca, la colpa dovrebbe essere uguale. E invece l'omicidio è più grave del tentato omicidio, questo è palese. Ma in virtù di cosa lo è?

È possibile mantenere PCo se ipotizzo che davvero il giudizio dipende solo dalle intenzioni, ma che a Tre e Quattro sono assegnate intenzioni diverse. L'intenzione è uno stato mentale, è privato e nessuno, tranne chi vive quello stato, ne ha accesso. Ciò che io posso fare, allora, è stimare l'intenzione in base al comportamento. Il miglior indizio del fatto che uno ha l'intenzione di fare qualcosa è il suo farla (Latus 2000). Al che, è naturale spiegarsi il fatto che Quattro sia riuscito nel suo intento, mentre Tre abbia fallito, ipotizzando che Quattro avesse una intenzione più forte e seria, che fosse più determinato, e quindi anche più colpevole. Questo può essere esteso anche al primo esempio: il guidatore Due, proprio perché ha ucciso, deve essere più negligente di Uno, e quindi più colpevole.

Posso obiettare: Tre avrebbe potuto uccidere il Presidente, è stata solo (s)fortuna che non ci sia riuscito. Questo è un controfattuale. Io, fossi nato nella Germania di Hitler, sarei stato un nazista? Come con l'intenzione, il più forte indizio in merito a ciò che avrei fatto è ciò che sto concretamente facendo oggi (Richards 1986): quindi, anche in tal caso, mi spiego il fatto che Tre abbia fallito con l'idea che avrebbe fallito in ogni caso, mentre Quattro avrebbe comunque trovato il modo di riuscire nei suoi intenti, e quindi sarebbe stato sempre più colpevole.

È una soluzione comoda perché esclude a priori la sorte: si giudica sempre ipotizzando che non vi siano elementi casuali. Epperò non mi sembra di giudicare così. Immagino: io conosco davvero le intenzioni di Tre e Quattro perché so che entrambi hanno scritto un programma di più di mille pagine nel quale ingiuriano il Presidente e spiegano ogni dettaglio del loro piano omicida; ho inoltre delle registrazioni degli attentati che mostrano come Tre e Quattro abbiano puntato la pistola esattamente nello stesso modo, e che il proiettile di Tre sia stato deviato solo grazie ad una improvvisa folata di vento. Quindi io so che, non fosse stato per il caso, Tre avrebbe davvero ucciso il Presidente; così è quasi impossibile sostenere che Tre e Quattro avessero intenzioni diverse. Eppure Quattro rimane più colpevole di Tre.

Faccio allora un'ipotesi diversa, sempre mantenendo PCo: Quattro è più colpevole di Tre perché ha infranto più principi. Poiché entrambi hanno tentato di compiere un omicidio, entrambi hanno infranto il principio «non mettere in pericolo gli altri», ma solo Quattro, essendo riuscito nell'intento, ha anche infranto il principio «non uccidere». Questo implica esternalismo: gli stati interni di Quattro sono identici a quelli di Tre, e lo stesso vale per il comportamento, ma i due hanno agito in contesti leggermente diversi, contesti tali da portare Quattro a infrangere quel principio in più.

È logico, «la Legge non ammette ignoranza»: i miei stati mentali, le mie intenzioni, sono irrilevanti; se lungo una strada c'è limite di velocità a cinquanta, affinché io infranga la Legge è sufficiente che io corra a sessanta, il mio saperlo o volerlo è indifferente. Il *dovere* «preserva la salute altrui» è infranto da Tre e Quattro al momento stesso dello sparo, ma il *diritto* del Presidente al «non essere ucciso» è infranto solo dopo, e solo da Quattro, in virtù del fatto che il contesto della sua azione ha previsto una diversa evoluzione del mondo, cioè diverse conseguenze.

Questa è una possibilità, ma: è ancora deontologia? Nell'etica kantiana il bene sta nell'agire per senso del dovere verso l'imperativo categorico, e l'imperativo categorico, che incarna sia diritti che

doveri, è rispettato o infranto nel momento stesso dell'azione. Quindi è da rifiutare l'idea che un certo diritto possa essere infranto diverso tempo dopo l'inizio dell'azione. Può essere allora che Quattro abbia infranto quel dovere in più al momento dell'azione? Ovviamente no: è insensato affermare che Quattro abbia infranto il diritto alla vita del Presidente al momento dello sparo e non al momento dell'uccisione, è insensato perché avrebbe potuto benissimo non morire.

Queste soluzioni che cercano di mantenere PCo non funzionano; quindi, se voglio formulare una deontologia compatibile con le intuizioni legate ai casi di sorte morale, devo capire in che contesti non è valido.

I casi di Tre e Quattro hanno come unica differenza una conseguenza. La deontologia non valuta le conseguenze *in sé*, ma è costretta a riconoscere una connessione tra la *responsabilità dell'azione* e la *responsabilità delle conseguenze*. Immagino che Quattro si opponga alla sua condanna davanti al giudice: «ma come a lui dieci anni e a me l'ergastolo? Non vede che io sono responsabile tanto quanto lui?» Cosa direi? Il fatto che Quattro affermi di avere tante colpe quanto Tre, foss'anche vero, lo rende un uomo ancora peggiore: il non prendersi la responsabilità delle conseguenze lo porta ad infrangere ancora più doveri. Ma quando devo, esattamente, prendere questa responsabilità?

Terzo esempio. Devo fare una consegna col furgone, mi metto alla guida, parcheggio, scarico i pacchi, una folata di bora potentissima rovescia il furgone su un vecchietto che stava passeggiando per di lì. Questa tragedia non sarebbe avvenuta se non mi fossi messo alla guida; ciò nonostante, non sono colpevole, perché:

1. ho rispettato tutti i miei doveri, non ho infranto alcun principio, regola, legge;
2. il rovesciarsi del furgone è indipendente dalle mie azioni (perché la raffica vi sarebbe stata comunque);
3. la presenza del vecchietto è indipendente dal mio controllo.

Finché si dà il punto uno, può non darsi il due, o il tre, o anche entrambi: se rimango all'interno della deontologia comunque non sono colpevole perché ho fatto tutto quello che dovevo fare. Adesso immagino che si diano il due e il tre, ma non l'uno: sono ubriaco, mi metto alla guida del furgone, vado a fare la consegna, scendo, ma la folata rovescia il furgone e uccide il vecchio. Per un certo consequenzialismo potrei essere colpevole perché se non mi fossi messo alla guida il vecchietto sarebbe ancora vivo, ma questo è assurdo: è vero che non ho rispettato un dovere, una legge, ma la folata rimane del tutto scollegata dal mio agire, non posso essere colpevole. Sembra così che anche il due blocchi la responsabilità delle conseguenze: se io sono responsabile della presenza del vecchietto (ad esempio, perché lo chiamo e gli dico di scendere che sto arrivando col pacco), comunque, se a rovesciare il furgone è una folata imprevedibile, non sono responsabile della sua morte.

Allora, poiché devo verificare quand'è che PCo non è valido, mantengo solo il punto tre, e immagino un caso in cui né l'uno né il due si danno: sono ubriaco, mi metto alla guida del furgone, sbando, il furgone si rovescia lì dove sta passando il vecchietto. Ho infranto dei doveri, e in più il rovesciarsi del furgone è dipendente dalle mie azioni in quanto sono io che ho sbandato; la presenza del vecchio nel punto dell'incidente rimane del tutto al di fuori del mio controllo, eppure sono colpevole della sua morte.

Ecco la connessione tra il mio agire e le conseguenze, il *principio della responsabilità*:

PCR: se agisci venendo meno ai tuoi doveri (usando gli altri come mezzi o in un modo che non accettabile come legge universale) accetta le conseguenze dipendenti dal tuo agire *come fossero* sotto il tuo controllo,

e quindi, corollario,

PCR*: comportati come se ne avessi colpa, cioè accetta la punizione.

PCo rimane valido, ma il principio della responsabilità lo sospende.

Se supererogo acquisisco nuovi doveri

La punizione non discende direttamente dall'infrazione delle regole, dal male commesso, ma può essere *sbloccata* dal fatto che quel male abbia certe conseguenze.

Se io rispetto tutti i principi, allora, anche se le mie azioni hanno conseguenze indesiderabili, io resto innocente; ma se infrango un principio perdo questa innocenza e, in un certo senso, divento colpevole anche di ciò di cui non dovrei essere colpevole. Questo è in linea con molti sistemi legali: un ubriaco può essere punito per omicidio stradale non perché ha controllo sugli eventi, ma perché quando si mette alla guida deve accettare responsabilità sulle conseguenze del suo agire qualsiasi esse siano, se è fortunato meglio per lui ma se è sfortunato deve accettare il carcere.

Il principio della responsabilità è così interpretabile come un caso di *supererogazione*: per non peggiorare la situazione etica sono disposto a riconoscere una colpa che non dovrei avere, ad accettare una punizione più dura di quanto sarebbe logico, faccio qualcosa al di là di quel che sarebbe normalmente richiesto; non può quindi costituire un dovere in senso proprio, è una regola che va al di là dei doveri.

La supererogazione non è del tutto estranea alla deontologia d'ispirazione kantiana: per esempio, Kant nega che il fare la carità sia un obbligo morale assoluto discendente dall'imperativo categorico, ma, come imperativo ipotetico, è comunque un bene che va al di là dei doveri universali. Parfit (1982), però, nota che la deontologia può così implicare casi difficilmente accettabili: io ho il dovere di salvare una vita se posso farlo, se vedo un uomo in arresto cardiaco ho il dovere di fargli la *CPR*; questo vuol dire anche che, se vedo una casa in fiamme con due bambini bloccati dentro, io ho il dovere di sfondare la porta e affrontare l'incendio anche a costo della vita? Certe deontologie rispondono di sì, ma è assurdo: io ho il dovere di chiamare i pompieri, non posso avere anche il dovere di dare la vita.

Questo è un problema per le etiche deontologiche: da un lato, se io ho il dovere di gettarmi tra le fiamme, i miei doveri sono così alti ed eccezionali che è impossibile supererogare, quando è invece intuitivo che un uomo possa andare al di là dei suoi doveri; dall'altro, se il bene sta nel fare il proprio dovere, e non è proprio dovere buttarsi tra le fiamme, perché il farlo dovrebbe essere una supererogazione, perché dovrebbe essere un bene? Anche l'utilitarismo classico ha difficoltà con questo punto, chiedendo di scegliere il massimo bene per il più alto numero di persone rende impossibile supererogare; ma se la deontologia non rende merito di tali intuizioni perde uno dei motivi per favorire il suo approccio a quello consequenzialista.

Immagino la casa in fiamme con due bambini. È vero: ho il dovere di salvare chi posso ma non anche *a rischio della vita*. Che sia così lo si nota facilmente controllando la tipologia del contesto: sia un mondo nel quale tutti si gettano tra le fiamme, sia un mondo nel quale nessun non-pompieri si getta (ma chiama i soccorsi), è razionalmente accettabile, quindi gettarsi non può essere un obbligo.

Cosa succede, però, se entro per salvare uno dei due bambini? È legittimo per me non tornare dentro a salvare anche l'altro? Se io, pur potendolo fare, scelgo ad arbitrio di non salvare uno dei due, sto peggiorando la condizione etica del mondo: infatti, anche se non ho il dovere di sacrificarmi, l'imperativo categorico mi chiede di essere giusto, equo, di non fare eccezioni. Questo ha un effetto sorprendente: io non ho il dovere di salvare né uno né l'altro; ma poiché i doveri funzionano a sistema, se mi butto tra le fiamme per uno (e ho ancora abbastanza energie), tale azione, interagendo con il dovere all'equità, genera il dovere di buttarmi anche per l'altro (cfr. Wessels 2015). Se dono ad una associazione benefica, la quale con cento euro può salvare una persona ma con centoventi può salvarne due, io non ho il dovere di donare alcunché, ma avendo il dovere di essere razionale, donando acquisisco il dovere di donare centoventi e non meno (si veda Horton 2017, Sinclair 2018).

Cosa dovrei dire di un pompiere che, avendo salvato un uomo dall'alluvione, rinfacciasse «non era mio dovere farlo, ora dovrete portarmi sul palmo di mano»? Costui non è un eroe, l'eroe ha per sua scelta più doveri di una persona normale. Questo *eroismo*, questa *super-responsabilità*, è ciò che costituisce la supererogazione in deontologia:

SR: chi va oltre i suoi doveri una volta sceglie di prendere su di sé doveri che in precedenza non aveva, e quindi deve comportarsi comunque come se agisse sempre e solo per dovere.

Sorge ora un ultimo problema per la supererogazione. Finora ho immaginato un contesto tale che, non avendo io al momento doveri da assolvere, mi si presenta l'occasione di compiere un'azione che, pur non essendo un dovere, è meritevole di essere compiuta. Ma cosa succede se io ho già un dovere da rispettare? Questo esempio viene dall'eticista Kamm (1996):

- ho il dovere di arrivare in orario ad un appuntamento, ma ho l'opportunità di salvare due bambini da un incendio (supererogazione);
- poiché salvare i bambini non è un dovere, salvarli è legittimo tanto quanto il non fare niente;
- ma essendo una supererogazione, salvarli è legittimo almeno tanto quanto l'arrivare in orario all'appuntamento;
- eppure, essendo un dovere, arrivare in orario all'appuntamento è più legittimo del non fare niente (che anzi è illegittimo).

Questo suona assurdo: se salvarli ha un valore pari al non fare niente e all'arrivare in orario, anche non fare niente e arrivare in orario avranno lo stesso valore. Nel sistema esposto, però, non funziona così: io ho il dovere di arrivare in orario, quindi non posso starmene immobile; se però mi fermo per salvare i bambini, e salvandoli diventa impossibile arrivare in orario, per SR e il *se devo allora posso*, acquisisco il dovere di salvare i bambini e perdo il dovere di arrivare in orario; perciò, i valori cambiano, e non posso confrontare doveri e supererogazioni nel modo statico dell'esempio.

Rimane un dubbio: se io devo fare il mio dovere, ma una supererogazione è un'azione buona che, pur non essendo un dovere, nel momento in cui sto per compierla diventa tale, come faccio a

riconoscere queste azioni, a distinguerle da quelle che sono ugualmente al di fuori del dovere ma che sono male? Questo lo dirò più avanti.

Un dovere può avere effetti sia positivi che negativi, ma non solo negativi: il dilemma del carrello

Negli ultimi anni si parla tanto delle auto a guida autonoma, auto capaci di portare in autonomia il loro proprietario da casa al lavoro e ovunque voglia andare. Queste macchine possono trovarsi in situazioni complesse anche da un punto di vista etico, e devono essere in grado di reagire adeguatamente.

Immagino di ordinare alla mia auto a guida autonoma di portarmi sul monte Pizzoc, sto rispondendo alle mail mentre l'auto affronta un tornante, sbuca dal nulla un veicolo che procede contromano ad alta velocità, ha a bordo un padre e tre bambini. L'auto calcola che ci sono solo due possibilità d'azione: o si getta nella scarpata, salvando la famiglia nell'altra auto ma causando la mia morte; o accetta di fare un frontale, salvando me con un sistema di airbag ma uccidendo la famiglia nell'altra macchina. Il programmatore di questa auto a guida autonoma deve predisporla a compiere, in un caso come questo, quale scelta? (si veda JafariNaimi 2018)

Utilitarismo e deontologia sembrano essere spesso sulla stessa lunghezza d'onda: pur in modi diversi arrivano a identificare lo stesso bene e lo stesso male. In questo caso non è così: per un utilitarista, totalmente abnegato verso i propri interessi, l'auto dovrebbe buttarsi giù dalla scarpata perché la morte di uno è meglio della morte di quattro; per un eticista kantiano, invece, il produttore dell'auto, al momento del contratto, ha preso su di sé il dovere di preservare (non uccidere!) l'acquirente, e quindi è legittimo che l'auto scelga il frontale. Quest'ultima sarà, credo, la scelta della maggior parte dei Paesi: l'azienda ha un legame unico con il cliente, diverso dal legame che ha con qualsiasi altro uomo, un legame tale da avere il dovere di proteggerlo anche se ciò causa la morte di altre persone.

Questo che ho esposto è un esempio di *problema del carrello ferroviario*, il più influente tipo di dilemma etico. Analizzerò ora un gran numero di altre versioni.

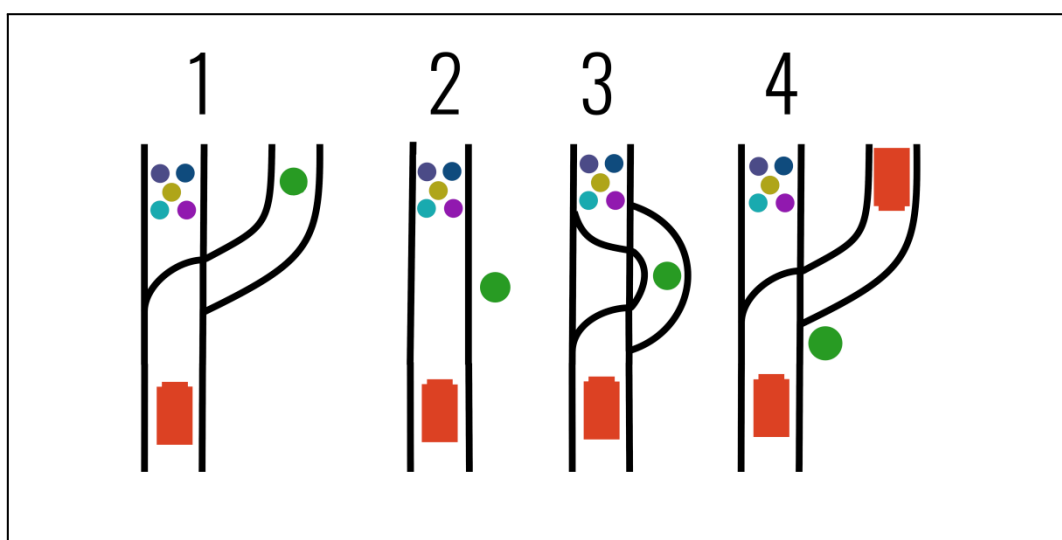


Figura 9

Prima versione. Un carrello ferroviario avanza a gran velocità, sui binari ci sono cinque persone che verranno uccise se non faccio niente. Proprio di fianco ho una leva di cambio binario, mi permette di deviare il carrello su un altro binario dove però c'è una persona, che verrà uccisa. È etico tirare la leva? (cfr. Foot 1967)

Seconda versione. Stesso carrello che avanza a gran velocità verso un gruppo di cinque persone; di fronte a me c'è un uomo di duecento chili, potrei dargli una spinta per farlo cadere sui binari e fare così in modo che il carrello si schianti contro di lui e si fermi, uccidendolo, ma salvando gli altri cinque. È etico dare una spinta all'uomo? Alternativa: l'uomo si trova su un ponte sopra il binario, tirando la leva aprirò una botola e lo farò cadere davanti al carrello, uccidendo lui ma salvando i cinque: è etico tirare la leva? (Edmonds 2014)

Terza versione. Ci sono sempre due binari, uno con cinque uomini e uno con un solo uomo, ma questo secondo binario è curvo e si ricongiunge al primo proprio davanti ai cinque uomini. Se non faccio niente, il carrello ucciderà i cinque; se tiro la leva, ucciderà l'uomo sul secondo binario, ma se costui non vi fosse stato il carrello avrebbe continuato il suo tragitto uccidendo comunque i cinque. Che faccio, tiro? (Thomson 1985)

Quarta versione. Il carrello si sta dirigendo verso i cinque uomini, tirando la leva posso sbloccare un secondo carrello vuoto che andrà a scontrarsi con il primo, fermandolo e salvando i cinque uomini; questo, però, farà deragliare il carrello e lo farà cadere giù per la collina fino a rotolare proprio addosso ad un uomo che sta riposando su un'amaca nel suo giardino, uccidendolo. Tiro? (Unger 1992)

In tutti questi casi gli utilitaristi scelgono di sacrificare una vita per salvarne cinque. Per la deontologia la questione è più complessa. Secondo alcuni filosofi, il fatto che sia necessario tirare la leva (o gettare l'uomo grasso sul binario) è così evidente che la deontologia, per poter essere una teoria sensata, deve dimostrare di poter accettare questa soluzione (e.g. Greene 2013); secondo altri, invece, tirare la leva è sbagliato, ed è l'utilitarismo a dover dimostrare di saper dare valore ad ogni singola vita (e.g. Thomson 2008).

Sono banale: inizio con la prima versione. L'imperativo categorico mi dà il dovere di soccorrere le persone in pericolo (un mondo in cui tutti se ne vanno senza aiutare è inaccettabile). Il mio tirare la leva, però, mette l'uomo del secondo binario in pericolo, e verso di lui avrò gli stessi doveri che prima avevo verso gli altri cinque; in più, in tal caso sono io la causa del pericolo: sembra, quindi, che io non debba tirare la leva.

C'è però un principio che può permettere alla deontologia kantiana di accettare il cambio di binario, la *dottrina del duplice effetto*:

DDE: se un'azione

- è neutra in quanto ha sia un effetto positivo che uno negativo di proporzionata gravità,
- ed è motivata da buone intenzioni, senza desiderio di fare un male (pur inevitabile),
- e i due effetti sono indipendenti, cioè nessuno è un mezzo per ottenere l'altro,
- né è possibile evitare o bloccare l'effetto negativo pur ottenendo il positivo,

allora la tale azione è comunque legittima (d'Aquino 1995-2013 II-II Q.64 Art.7).

Se si accetta ciò, è vero che tirando la leva causo la morte di una persona, ma causo anche la salvezza di cinque, e la salvezza di questi cinque è indipendente dalla morte dell'altra persona (fosse morta oppure no, loro sarebbero salvi comunque), né posso fare qualcosa per salvarla, quindi è una scelta legittima.

Tuttavia, c'è un altro principio che spinge in direzione opposta, la *distinzione tra commettere e permettere*:

DCP: il commettere un male in modo attivo è più grave del permetterlo assistendovi passivamente

Ad esempio: chi osserva una rapina e non fa alcunché per impedirla fa un male, ma non grande come chi la sta commettendo.

Ci sono numerose critiche a DCP, ma è intuitivo che il principio valga almeno in alcuni contesti. Immagino due casi: nel primo, affogo una persona; nel secondo voglio affogarla, ma quando sto per farlo, le prende un infarto e decido invece di non soccorrerla (Rachels 1975). In un contesto con intenzioni malvagie, sia commettere che permettere sono ugualmente gravi. Di contro, però, immagino altri due casi: nel primo, viaggio di notte a fari spenti, non vedo un pedone, lo travolgo e uccido; nel secondo, un pedone sul marciapiede ha un infarto, io viaggiando a fari spenti non lo vedo, e così non lo soccorro. Nel primo caso, anche se ho ucciso in modo involontario, avendo infranto un principio, sono perseguibile; nel secondo, è vero che ho una parte di colpa (perché se avessi avuto i fari accesi l'avrei visto e soccorso) ma comunque non sono perseguibile. Il principio pare dunque valere se si tengono in considerazione le intenzioni. È possibile che in futuro si dimostri che DCP è in realtà solo il riflesso di un altro diverso principio (una distinzione tra azione e inazione per esempio, come in Quinn 1989) ma questo è irrilevante nel presente contesto.

Se si accetta questa distinzione, allora, nel caso del carrello ferroviario se io non faccio niente permetto (non causo) la morte di cinque persone, ma se tiro la leva causo (e non permetto solo) la morte di un uomo, e questo è più grave, quindi non devo farlo. Posso obiettare che, poiché la morte di quell'uomo è indipendente dalla salvezza dei cinque, anche la sua morte non è causata ma solo permessa, sto solo *omettendo* di spostare la leva una seconda volta per salvarlo. Questo però suona come un sofisma razionalizzante: io ho un ruolo nella sua morte, posso prevedere che morirà e che se non avessi tirato la leva non sarebbe morto, non sto solo accettando un evento indipendente dal mio agire.

Dato che i due principi esposti sono entrati in contraddizione, vado a ripescare anche il principio $D \rightarrow P$, *se devo allora posso*: è infatti evidente che io non posso togliere i cinque uomini dal pericolo senza mettere in pericolo l'uomo del secondo binario, e viceversa, quindi non posso creare una condizione in cui nessuno è in pericolo, e quindi non ho neanche il dovere di farlo, cioè i due principi non possono generare entrambi obblighi. Quale lo fa?

Riprendendo i principi dell'imperativo categorico, senza dubbio non sto usando l'uomo del secondo binario come un mezzo, la sua presenza è indifferente alla salvezza dei cinque. Ha lui forse un diritto ad essere salvato? Direi di no: se io sono un riccastro ho il dovere morale di fare la carità, ma nessun singolo mendicante ha, di per sé, il diritto a riceverla in quanto non posso (e quindi non devo) farla *a tutti*, il fare la carità solo ad alcuni non lede il diritto di chi non la riceve. Quindi, al mio dovere di salvare non corrisponde un diritto ad essere salvati, i diritti di tutti i coinvolti sono comunque rispettati e l'imperativo categorico è soddisfatto. Se è così, in questa deontologia è legittimo tirare la leva.

Ciò nonostante, sarà vero che i principi D→P e DDE consentono di tirare la leva, ma non la indicano come la scelta migliore: sia tirare la leva che non tirarla rispettano tutte le condizioni del duplice effetto; inoltre, come l'uomo sul secondo binario non ha diritto ad essere salvato, così neanche gli altri cinque lo hanno, ed è quindi legittimo lasciarli investire. Posto ciò, DCP diventa determinante, e conclude di non tirare la leva: questa è la risposta comunemente accettata dai sostenitori della deontologia (Bourget & Chalmers 2014).

Si potrebbe sostenere, dato che i principi entrano in contraddizione, che bisogna mantenere solo uno dei tre. Non intendo però così questa etica: tutti i principi visti, se presi in isolamento, potranno a conclusioni inaccettabili, ma se considerati tutti insieme e soppesati, allora portano alle giuste decisioni.

Rifletto ora sulle restanti tre versioni. La seconda evidenzia i diritti: se io butto l'uomo da duecento chili sul binario infrango il principio d'umanità dell'imperativo categorico in quanto userei il suo corpo come mezzo per raggiungere il fine di salvare i cinque, quindi la deontologia kantiana rifiuta senza dubbio questa scelta. Cosa dire della variazione nella quale l'uomo si trova su un ponte, su una botola proprio sopra i binari? (Greene et al. 2009) La base non cambia: tirare la leva viola il principio di umanità e quindi non posso farlo. Ciò nonostante, l'intuizione etica comune è che il far cadere l'uomo tirando una leva sia meno grave dello spingerlo, forse perché si depersonalizza il contatto.

Nella terza versione il fatto che il binario sia curvo impedisce l'applicazione della dottrina del doppio effetto in quanto la salvezza dei cinque è ora dipendente dalla presenza sul binario dell'uomo, se non ci fosse o riuscisse a scappare i cinque morirebbero comunque. Perciò, a differenza della prima versione, tirare la leva in questo caso violerebbe il principio di umanità, ed è senza dubbio inaccettabile.

La quarta versione si differenzia per il suo coinvolgere un uomo esterno al problema: il fatto che l'uomo non sia sul binario, ma in un giardino lontano, non cambia nulla per la deontologia, tirare la leva rimane inaccettabile; ciò nonostante, l'intuizione etica è che questo caso sia più grave di quello della prima versione proprio perché l'uomo non era originariamente coinvolto nel sistema dei binari.

La domanda che questi casi mi sollevano è: quali sono gli elementi *eticamente rilevanti*? La seconda versione chiede di distinguere la caduta dell'uomo sui binari per una spinta dalla caduta causata da un diverso meccanismo indiretto (foss'anche una *macchina di Rube Goldberg*); la terza versione chiede di distinguere tra dipendenza e indipendenza del sacrificio dell'uomo; la quarta versione chiede di distinguere tra un uomo presente sui binari e un uomo più lontano. Le risposte etiche delle persone comuni variano in base a questi elementi, e variano anche in base all'ordine con il quale vengono presentate le quattro versioni, un po' come il verdetto di un giudice può cambiare in base ai *precedenti*.

Un ulteriore elemento di questo tipo è l'identità delle persone coinvolte. Chi sono le persone sui binari? In una situazione concitata l'etica non è sospesa, devo poter decidere bene anche se non riesco a identificarli. Però potrei anche riuscirci, e potrei scoprire che l'uomo del secondo binario è il *mahatma* Gandhi, mentre i cinque sul primo binario sono tutti assassini la cui condanna a morte era prevista per domani: c'è ora una più grande pressione a non toccare la leva. Di contro, potrei scoprire che i cinque del primo binario sono dei disgraziati, e l'uomo sul secondo binario è il boss mafioso che li ha fatti incatenare lì per ucciderli: ora sarei molto più propenso a tirare la leva (cfr. Scruton 2018).

Mi chiedo: si tratta di *bias*, elementi irrazionali che influenzano il giudizio ma non dovrebbero, o sono invece questi elementi importanti che le teorie etiche devono prendere in considerazione se non vogliono fraintendere il concetto di *moralità*? Cioè, le intuizioni etiche devono avere un ruolo nel giudizio e nell'individuazione della migliore teoria normativa, oppure sono troppo inaffidabili?

Almeno l'identità degli individui è pare essere un fattore importante. L'utilitarismo lo nega e non la considera, uno vale uno. La deontologia è più aperta: da un lato, ogni individuo ha diritto alla vita, il fatto che l'uomo del secondo binario sia un boss non può arrivare a giustificare la sua uccisione; dall'altro, il fatto che sia un boss, e quindi abbia infranto dei principi, dà alla società il dovere di punirlo e fa decadere certi suoi diritti, di conseguenza potrebbe essere legittimo lasciarlo investire senza soccorrerlo.

Curiosità finale: l'educatore Michael Stevens (2017), noto agli *youtuber* come Vsauce, ha inscenato dal vivo il problema del carrello; le persone coinvolte, incaricate di scegliere se cambiare binario oppure no, erano convinte che ciò a cui stavano assistendo fosse reale. La maggior parte ha scelto di non tirare la leva. Si può ipotizzare che le persone comuni applichino istintivamente una forma di deontologia più che di utilitarismo; d'altro canto, si può rispondere che le persone comuni non sono esempi di virtù morale eccezionale, quindi tale risultato è irrilevante.

Il dovere può essere ad personam: tre nuovi dilemmi

Primo dilemma. Adelchi e suo figlio Baodolino, l'amica Cunegonda e i suoi figli Desiderio ed Ermetruda stanno facendo un giro in motoscafo anche se tra loro Adelchi è l'unico che sa nuotare. Urtano uno scoglio, Adelchi e Cunegonda ricadono all'interno del motoscafo ma Baodolino, Desiderio ed Ermetruda sono sbalzati fuori; Baodolino è caduto lontano a destra, mentre Desiderio ed Ermetruda sono vicini a sinistra; Cunegonda non può buttarsi in mare perché affogherebbe, Adelchi è disposto a buttarsi ma, essendosi ferito, ha energie sufficienti per recuperare o suo figlio Baodolino, o i due figli di Cunegonda. Chi dovrebbe salvare? (cfr. Williams 1981)

Secondo dilemma. Entro in ospedale per dei controlli postoperatori. Arriva un dottore, mi spiega che purtroppo ci sono cinque pazienti in gravi condizioni: uno di loro ha bisogno di un trapianto di polmoni, un altro di un trapianto di cuore, due hanno bisogno di un rene, e l'ultimo di un fegato nuovo; ha visto le mie analisi, che sono tutte in regola, e mi chiede se, così per caso, non fossi disposto a suicidarmi per offrire i miei organi ai cinque pazienti in condizioni critiche. Se non lo faccio, moriranno cinque persone, ma se io mi suicido, alla morte di uno seguiranno cinque vite salve: cosa dovrei fare? (Thomson 1985)

Terzo dilemma. Poiché nessuno vuole suicidarsi per donare gli organi, viene creato un allevamento di esseri umani, un laboratorio dove, tramite un utero artificiale, un feto, progettato per nascere con danni cerebrali tali da non fargli mai prendere coscienza, viene fatto sviluppare con il solo scopo di poter poi usare i suoi organi. È morale o immorale farlo? (cfr. Gilman 2012)

Un utilitarista estremo non discrimina: bisogna agire per salvare più persone possibile (a patto che non sia certo che queste persone useranno la loro vita per fare del male) (cfr. Kagan 1989). La deontologia, invece, può riconoscere l'esistenza di rapporti speciali, come i legami di parentela, che creano quelli che chiamo *doveri ad personam*: Adelchi ha un legame unico con Baodolino, un legame che non ha con i figli di Cunegonda, e che gli dà il dovere di salvare prima lui; e io ho un legame

unico con me stesso che mi dà il dovere di proteggermi e preservarmi anche se ciò va a danno di altre persone.

Il terzo caso è più complesso, ma la deontologia tende a rifiutarlo. Quei feti hanno diritti *personali*? La discussione è infinita, ma il dilemma presuppone come criterio della personalità (cioè dell'essere *persona*) la coscienza; la domanda allora diventa: la personalità dipende dalla coscienza? Non pare essere così. Una società non ha coscienza, ma può avere dei diritti e dei doveri. Un uomo in stato comatoso potrebbe non acquisire mai più coscienza, eppure non perde immediatamente la personalità. Se io sparo ad un uomo addormentato, non lo privo della coscienza, che non ha, lo privo solo della possibilità di acquisirla in futuro (come si fa manipolando il feto), ma questo rimane omicidio. Posto ciò, non c'è motivo per il quale quei feti non debbano avere un insieme minimale di diritti: distinguere l'organismo dalla personalità, dire che ci sono certi organismi *vivi* che non sono persone (perché incapaci di coscienza), oltre che controintuitivo, è pericoloso perché strumentalizzabile per scopi razzisti. Se non si trova un altro motivo per negare il loro statuto personale, usarli per raccogliere organi va contro il principio di umanità, e quindi è inaccettabile.

Il dovere varia con le conseguenze intese: un problema generale

Sia pescando dalla tradizione, sia affidandomi ad intuizioni personali, ho esposto un tipo strutturato di deontologia kantiana: principio di universalizzabilità, di umanità, di controllo, di responsabilità, del duplice effetto, Regno dei Fini, se devo allora posso, quadripartizione dei casi di obbligo e divieto, super-responsabilità dell'eroismo, distinzione tra commettere e permettere, doveri ad *personam*, si combinano per creare un sistema in grado di affrontare molte situazioni diverse. Epperò resta una deontologia, e in quanto tale è sensibile alle obiezioni che attaccano questo tipo d'etica nell'insieme, cioè quelle che cercano di mostrare la necessità di riconoscere eccezioni ai doveri in base alle conseguenze.

Torturare è male: si usa una persona per ottenere a forza un'informazione, quindi viola il principio di umanità. E non è solo contro l'imperativo categorico, ma anche contro il suo imperativo ipotetico: non è un buon mezzo per ottenere informazioni veritiere in quanto motiva il torturato a mentire per far cessare il dolore e perdere tempo. Ciò nonostante, il dilemma etico che va a colpire le basi della deontologia è spesso esposto come un caso di tortura: farò finta, allora, che sia un mezzo efficiente per arrivare alla verità.

Un magistrato è stato minacciato di morte da un terrorista, costui è latitante ma si è riusciti a mettere le mani su sua moglie e i suoi figli. Costoro, pur essendo innocenti ed estranei alle vicende, hanno il *dovere ad personam* di proteggere il loro familiare, quindi non riveleranno la sua posizione; però è possibile torturarli per spingerli a rivelarla, e questo permetterà di prevenire l'attentato e salvare il magistrato. È legittimo torturarli?

La deontologia risponde di no: non posso usare persone come mezzi neanche per salvare delle vite. Occhi, allora rivedo l'esempio così: il terrorista ha rubato un ordigno nucleare che userà per uccidere milioni di persone, torturando la moglie è possibile conoscere la sua posizione, recuperare l'ordigno e salvare milioni di vite. Ora è legittimo torturarla? Dovrei davvero essere così testardo da dire «perisca pure il mondo intero, non verrò meno ai miei principi»? (Bagaric & Clarke 2007)

Questo tipo d'esempi tormentano la deontologia sin dai tempi di Kant. Lui era ossessionato dalla sincerità, credeva che mentire fosse inaccettabile in tutti i casi. Al che, immagino di dare ospitalità

ad un disgraziato in fuga, arriva da me il suo persecutore che mi ordina di dirgli dove lo sto nascondendo perché vuole ucciderlo: è davvero sbagliato mentire? I *giusti tra le nazioni* che hanno nascosto degli ebrei e hanno mentito alle autorità naziste che li stavano cercando hanno dunque fatto qualcosa di immorale? Kant risponde di sì: non sono io che devo mentire, è l'altro che non deve uccidere, io al massimo posso tacere (Kant 1797). Ma questo è assurdo.

Tali esempi sono problematici per questo: la deontologia basa il bene su doveri assoluti, se mentire è contro l'imperativo categorico non dovrebbe mai esserci una circostanza nella quale sia bene mentire; qui si mostra invece che è razionale riconoscere eccezioni ai doveri in virtù della positività delle conseguenze. Epperò la deontologia kantiana che ho formulato riesce a tenere conto delle conseguenze, per esempio deve farlo per poter individuare una contraddizione pragmatica nella volontà. Posso così notare che:

- il bene è tale che la sua aggiunta non può rendere peggiore la condizione etica del mondo, ma dire la verità ad un assassino e permettergli così di uccidere rende la situazione peggiore, quindi non può essere un bene;
- se io considero l'azione *nel contesto*, cioè non «mentire» ma «mentire *per* salvare un perseguitato», e applico la versione negativa dell'universalizzabilità chiedendomi «cosa accadrebbe se nessuno lo facesse mai?», vedo che in realtà si tratta di un obbligo;
- per il principio di responsabilità, se io vado contro il mio dovere devo accettare le conseguenze dirette dell'azione come sotto il mio controllo, ma se queste conseguenze sono del giusto tipo e positive, o almeno lo sono *oltre un certo livello*, allora, pur avendo commesso un male non facendo un mio dovere, ricevo dalle conseguenze anche un merito che può oltrepassare la colpa di quel male.

Immagino che il giudizio sul «mentire» e il giudizio sul «mentire *per* salvare un perseguitato» siano due diversi giudizi. Questo vuol dire che la regola «non mentire...» include l'eccezione, «...a meno che non sia per salvare». Quindi, prima l'imperativo categorico giudica negativamente il «mentire», e poi lo stesso imperativo categorico riconosce le conseguenze della regola «non mentire (neanche per salvare)» come razionalmente inaccettabili, quindi crea il dovere di venirvi meno per evitarle. Ma questo è impossibile per il principio *primum non nocere*, implicitamente accettato da Kant:

PNN: devo agire per il bene, e non posso fare il male neppure per prevenire che altri facciano del male in futuro

Cioè non posso venir meno al mio dovere neppure per evitare che altri vi vengano meno (cfr. Ippocrate 1923). Tale principio è irrinunciabile: se lo si negasse, allora, posto che probabilmente in futuro vi saranno crimini, omicidi, guerre innumerevoli, la scelta più etica sarebbe sterminare oggi tutta l'umanità, un gran male che però previene ogni possibile delitto futuro chiudendo con un bilancio positivo. Questo è inaccettabile, quindi PNN va mantenuto: le regole non possono avere eccezioni, né può esservi una valutazione delle conseguenze contraria alla valutazione dell'azione.

Mentire per ingannare è male, mentire per salvare è bene: è impossibile valutare separatamente l'azione e il suo scopo, cioè le sue conseguenze intese, quindi la differenza tra i due casi va spiegata non col fatto che uno dei due è un'eccezione, ma che corrispondono a due regole diverse. Considerare l'azione *nel contesto* vuol dire considerare l'azione insieme alle sue conseguenze intese, ovvero alla sua motivazione di primo livello (che fa da base a quella di secondo livello, la buona volontà).

Il problema ora è: com'è possibile individuare gli scopi legittimi, cioè in cosa consiste il loro dare merito, la loro accettabilità razionale, il miglioramento che apportano al mondo, e il loro fare da base alla buona volontà (cioè costituire un dovere)? Certe azioni potrebbero essere tali da avere sempre uno stesso giudizio qualsiasi sia lo scopo, ma ciò non può essere vero per tutte; questa deontologia deve quindi essere integrata con un criterio che permetta di dire quali scopi sono legittimi e per quali azioni.

La tesi dell'etica della virtù 1

L'idea della virtù ha dominato la filosofia morale per tutta l'epoca antica e medievale, da Aristotele (1999) a Tommaso; poi, nell'epoca moderna, il concetto ha perso via via importanza. Col secolo scorso i filosofi hanno iniziato a dirsi insoddisfatti delle teorie etiche moderne, molti hanno creduto che il progetto moderno di fondare un'etica sulla ragione (come in Kant) o addirittura sul calcolo (come in Bentham) fosse rovinosamente fallito; quindi non restava che guardarsi indietro, al modo in cui si filosofava prima (Stocker 1976).

Wittgenstein (2009, cfr. 1967) scriveva che, anche esistesse un libro contenente tutti i fatti, tale libro non presenterebbe alcuna affermazione di tipo etico: una scienza della morale è quindi impossibile, il bene non è qualcosa che si può scoprire con l'indagine. La sua allieva Anscombe (1958) sosteneva che il fallimento delle etiche moderne stesse nel fatto che non hanno fondato la morale in un legislatore intrinsecamente morale come Dio. Ma allora cosa dovrebbe credere chi non ha fede? Se Kant e Mill non funzionano più bisogna tornare ad Aristotele: ecco l'etica della virtù; la virtù, come bene intrinseco, permette di dare un fondamento laico, oggettivo ma non assoluto, alla morale. Philippa Foot, filosofa atea, fu convinta da Anscombe della necessità di dare un fondamento oggettivo alla morale: Foot (2008) è così diventata una delle fondatrici della moderna etica delle virtù; la segue Alasdair MacIntyre, autore di *Dopo la Virtù* (2007), che con quest'opera ha dato inizio alla *svolta aretaica*, il rinnovato e generale interesse per la l'etica aristotelica.

Cos'è quindi la virtù? È un tratto personale, un modo di essere, una proprietà che rende buoni; non riguarda solo il comportamento, ma anche gli interessi, il modo di pensare e di percepire, le emozioni, i desideri; è una tendenza generalissima che si traduce (ma non si esaurisce) in un modo abituale di agire. È magnanimo non chi agisce magnanimamente, ma chi ha emozioni e interessi magnanimi, sceglie amici magnanimi, apprezza le cose che esprimono magnanimità: il magnanimo ha una vita magnanima in tutti i suoi aspetti (Hursthouse & Pettigrove 2018). Una persona virtuosa non andrà a trovare l'amico malato «perché è virtuoso farlo», ma perché ha motivazioni virtuose, cioè perché riconosce il valore intrinseco del visitare gli amici.

Dare valore solo alle conseguenze vuol dire agire in modo cinico e malvagio solo perché questo potrebbe avere conseguenze piacevoli; dare valore solo alle azioni vuol dire agire in modo cieco e ipocrita non piegandosi neanche davanti alle catastrofi. Ma la virtù non sta nell'azione, non sta nelle conseguenze, sta tra le due cose: la virtù sta nei processi di scelta. L'uomo virtuoso non è obbligato né ad agire sempre in uno stesso modo qualsiasi siano le circostanze, né a puntare sempre ad un unico scopo; quindi non approverà azioni buone con conseguenze catastrofiche, né azioni malvagie con buone conseguenze, bensì si adatterà al contesto scegliendo azioni virtuose dirette verso conseguenze virtuose.

L'etica della virtù può così essere interpretata come una teoria più generale di consequenzialismo e deontologia, tanto da inglobarle: sia «agire per aumentare la felicità» sia «rispettare i doveri» possono essere virtù. I sostenitori di consequenzialismo e deontologia, da parte loro, hanno tentato di inglobare la virtù nei loro sistemi: un utilitarista può dire che un'azione buona ha la conseguenza di sviluppare e manifestare virtù in quanto la virtù è sufficiente alla felicità, un sostenitore della deontologia può dire che le virtù formano una classe di doveri («sii magnanimo», «sii casto...»); quindi, pur incompatibili tra loro, consequenzialismo e deontologia sono entrambe compatibili con l'etica della virtù. Ciò che distingue questa teoria è il suo concentrarsi non su caratteristiche dell'azione, ma su caratteristiche di chi agisce: se un'azione è attuata da un soggetto virtuoso allora è buona, se non l'avesse scelta allora non sarebbe virtuoso.

La tesi dell'etica della virtù 2: il benessere

Il concetto di virtù è più complesso di quello di felicità o dovere: questi due sono intuitivi, ci si può permettere di non definirli; cosa sia esattamente la virtù, invece, non l'ho ancora capito, sono stato troppo vago. Taglio la testa al toro: quali sono le virtù?

Per Aristotele (1999), la virtù sta nel mezzo: il coraggio è la mezza via tra codardia e incoscienza, la temperanza la mezza via tra insensibilità e licenziosità, e via così ad individuare anche liberalità, magnificenza, magnanimità, mansuetudine e varie altre. San Tommaso (sulla base della Repubblica di Platone e 1 Cor 13) propone le classiche virtù cardinali e teologali: prudenza, giustizia, forza, temperanza, e fede, speranza, carità. Benjamin Franklin, polimata e padre fondatore degli USA, ogni giorno registrava in una tabella i progressi e le mancanze fatte relativamente alle virtù che voleva sviluppare: ordine, silenzio, risolutezza, operosità, moderazione, castità, pulizia, umiltà, tranquillità e altre (Ford 1921).

Le virtù sono tante, milioni di milioni: chi tra questi pensatori dice la verità? Sembra impossibile determinarlo, e questo fa sorgere un'obiezione: se il bene è oggettivo, come può fondarsi sulla virtù, la quale varia con la cultura del luogo e del tempo? (Solomon 1988) Ma questo, per i teorici della virtù, non è un problema, anzi è un pregio: è normale che la comunità che deve praticare la virtù la ridefinisca di epoca in epoca, la virtù deve adattarsi alle varie circostanze, sono le altre teorie etiche che, pretendendo di individuare principi validi per tutti i tempi, principi distaccati dalla vita reale, risultano irrealistiche.

La miglior tassonomia delle virtù è offerta non dalla filosofia, ma dalla psicologia positiva. Mentre buona parte della psicologia tradizionale si dedica allo studio di casi patologici, la psicologia positiva si dedica all'eccellenza, alla felicità, al benessere, alla genialità; e per questo ha interesse nella virtù, che è appunto eccellenza. Gli psicologi Martin Seligman e Christopher Peterson hanno esposto in *Character Strengths and Virtues* (2004) la loro teoria, frutto di indagini empiriche transculturali volte ad isolare quei tratti che tutti, a prescindere dal loro background, possono riconoscere come positivi. Hanno individuato così sei gruppi di virtù:

- Sagacia (creatività, curiosità, apertura mentale, studiosità, lungimiranza, innovazione);
- Coraggio (audacia, persistenza, integrità, vitalità, spirito d'avventura);
- Umanità (amore, gentilezza, destrezza sociale);
- Giustizia (spirito civico, leadership, equità);
- Temperanza (clemenza, umiltà, prudenza, autocontrollo);
- Trascendenza (amore per la bellezza, gratitudine, speranza, humor, spiritualità).

Queste virtù non si limitano a disporre chi le possiede ad agire in certi modi, ma gli danno anche un obiettivo virtuoso: una virtù, per essere tale, deve puntare all'*eudaimonia*, il concetto greco traducibile come *benessere*, ma che corrisponde più precisamente al *senso della vita*, allo scopo che ogni persona ha in questo mondo, alla *vita felice* non di una felicità qualsiasi ma «di ciò che dovrebbe essere la felicità», alla vita vissuta così come va vissuta non per piacere o desiderio ma perché si riconosce il suo valore intrinseco.

Anche il benessere è uno dei temi studiati dalla psicologia positiva. Ci sono svariati fattori che correlano con esso: conoscendoli è possibile sapere come attuare queste virtù, è possibile sapere verso quale futuro si deve andare e come. Sempre Seligman (2017) ha proposto la teoria *PERMA* che individua cinque elementi del benessere:

- emozioni positive;
- impegno (lo stato di *flow* nel quale si è così concentrati su ciò che si sta facendo da non accorgersi più del mondo esterno, si veda Csikszentmihályi 1990);
- relazioni significative (sia romantiche che non);
- dare significato alla vita (tramite senso d'appartenenza a qualcosa di più grande di sé, come una religione, un ideale politico o altro);
- successo e senso di realizzazione.

Virtù e benessere *eudaimonico* formano un tutt'uno. Ogni virtù ha un suo campo di applicazione (per esempio, il coraggio riguarda le circostanze pericolose), in ogni campo la virtù richiede di agire in un certo modo per raggiungere un certo scopo (il coraggio difende i valori per superare la paura): così umanità, temperanza, trascendenza sono la via per sviluppare emozioni positive; sagacia, coraggio, giustizia favoriscono l'impegno; umanità e temperanza permettono di avere relazioni significative; eccetera (cfr. Swanton 2003).

Queste virtù possono essere intese come personali, come caratteristiche della persona; similmente l'*eudaimonia* può essere intesa come un certo modo di sentirsi individualmente. Queste interpretazioni però non sono sufficienti ad una teoria della virtù completa. Se interpreto l'*eudaimonia* come benessere, allora essa non dipende solo da fattori personali, ma soprattutto da fattori sociali: il livello di corruzione, il PIL pro capite, la libertà economica, il rispetto dei diritti umani, l'aspettativa di vita, l'indice di disoccupazione, il clima, la religione, sono tutti fattori che correlano con il benessere (Eger & Maridal 2015, Lucas & Lawless 2013, George et al. 2000). Le caratteristiche personali non sono sufficienti a rendere merito né della scelta dei mezzi, né della possibilità di raggiungere l'*eudaimonia*: il contesto in cui si nasce ha un effetto enorme.

Nagel (1991) presenta questo come un tipo di sorte morale: una persona virtuosa nata nella Germania nazista si trova in un contesto così profondamente corrotto che attuare la sua virtù personale diventa impossibile, e così finisce per comportarsi in modo corrotto a sua volta; ma, fosse nata negli USA, sarebbe stata una persona splendida. Questo richiede di estendere il concetto di virtù per inglobare anche *virtù sociali*: cioè l'azione buona non può essere frutto solo di un essere umano virtuoso, bensì deve esserlo di un essere umano virtuoso *in un contesto virtuoso*; per quanto la virtù del singolo possa brillare anche in un mondo marcio, comunque questo splendore personale ha dei presupposti esterni, come una giusta educazione in famiglia. Il concetto di virtù sociale è scarsamente analizzato nella filosofia contemporanea, ma è implicito nel pensiero greco e si trova già in Aristotele.

La tesi dell'etica della virtù 3: la saggezza

Per la virtù è facile spiegare i casi di sorte morale: tramite le virtù sociali accetta che la responsabilità dipenda da fattori esterni al controllo. Ha invece notevoli difficoltà a gestire i dilemmi etici come quello del carrello ferroviario: sia buttare l'uomo sulle rotaie sia non farlo può essere virtuoso a seconda delle circostanze e di quale virtù si vuole esprimere; quindi io cosa dovrei fare? Se getto l'uomo sui binari, dovrei essere punito oppure no? Suppongo esista una norma che dica di sì: questa è una norma giusta o ingiusta? Non si può dire, perché la virtù riguarda le persone, non le norme, non esiste una «norma coraggiosa». Ma allora come si può usare l'etica della virtù per arrivare a norme di comportamento?

La virtù non è della Legge ma del legislatore: sono buone le norme di un buon legislatore. Quindi, posto che il comportamento manifesta la norma, potrei scegliere se gettare l'uomo sui binari o no chiedendomi come agirebbe un ente virtuoso: l'etica della virtù è un'alternativa laica all'etica religiosa perché, come posso chiedermi «cosa farebbe Gesù?», così posso chiedermi «cosa farebbe Aristotele? Come agirebbe una persona perfettamente virtuosa?» Ma questa è una cattiva domanda: primo, perché io sono tutt'altro che perfettamente virtuoso e non potrò capire mai come agirebbe una persona che lo è (e, lo capissi, non riuscirei ad attuarlo); secondo, perché una persona perfettamente virtuosa deve agire tenendo in considerazione questo suo stato, tenendo in considerazione che ha già sviluppato ogni virtù, mentre io devo agire considerando quali virtù devo ancora sviluppare.

Non è dunque possibile risolvere certi dilemmi tramite norme, neanche potessi sapere cosa farebbe un super-virtuoso *osservatore ideale* (cfr. Firth 1952): l'etica della virtù non è interessata a formulare norme di comportamento, non vuole dirmi cosa dovrei fare in vari contesti, anzi afferma che chi giudica e decide solo tramite un'applicazione meccanica delle norme non è affatto virtuoso. Ciò che chiede di seguire nel giudizio è invece la *phronèsis*, la saggezza.

Cos'è la saggezza? In psicologia è la capacità di adattarsi a situazioni nuove e trovare soluzioni a problemi mai affrontati prima (cfr. Greene & Brown 2006). Si tratta però di una definizione errata: questa è l'*intelligenza fluida*, la quale raggiunge il picco verso i vent'anni e poi cala (Sternberg 1990, 2003); ma la *phronèsis* dei greci è esattamente il contrario, è una saggezza che per definizione un giovane non può avere, «ragazzo saggio» è un ossimoro.

La *phronèsis* sta nella capacità, davanti ad un problema, di immaginare diverse soluzioni possibili per poi confrontarle e soppesarle. Tale capacità presuppone una ricca esperienza di vita, richiede di variare frequentemente il proprio contesto, di fare *esperimenti* per vedere come diversi stili di vita portano a cambiare nel carattere e nel pensiero, come il proprio mondo si trasforma in funzione del comportamento: per questo è un qualcosa che un ragazzo non può mai possedere. Per i greci la saggezza richiede non solo conoscenza teorica astratta, ma anche abilità atletiche, oratorie, musicali; in generale richiede di sviluppare in modo armonico tutte le proprie abilità e potenzialità e poi coordinarle per realizzare gli scopi. Goethe era uomo di una saggezza olimpica perché aveva studiato molte cose diverse, aveva fatto molti lavori diversi, aveva sempre perseguito molti hobby diversi, e questo lo ha portato a capire il mondo, capire la mente, a creare cose, avere intuito, capacità d'autocritica, assenza di pregiudizi; in una frase, aveva vissuto tanto e visto tante cose che gli hanno permesso di raggiungere un invidiabile equilibrio.

La virtù, tra tutte le esperienze possibili, ne richiede una in particolare: l'esperienza di ciò che è bene (Hursthouse & Pettigrove 2018).

Il virtuoso non passa tutto il tempo ad autoanalizzarsi, al contrario guarda fuori di sé, abbandona l'ego e si lascia ispirare dagli atti di carità, dalla bellezza, dalle altre persone virtuose che lo circondano. È corretto che io mi chieda come agirebbe una persona perfettamente virtuosa, ed è ancora più corretto che io la osservi con attenzione, ma il bene non sta nel copiarla, non sta nell'agire come agirebbe: la *phronèsis* contempla il virtuoso per capire cos'è la virtù e poi sviluppare un modo personale di manifestarla e attuarla in funzione delle caratteristiche uniche proprie e del proprio contesto. Ogni uomo arriva all'*eudaimonia* per una strada diversa, ma l'*eudaimonia* non è un obiettivo, uno stato del mondo: ci sono più mondi possibili *eudaimonici* e ci sono più mezzi per arrivarvi, di modo che solo la *phronèsis* permette di adattarsi, scegliere tra alternative che in astratto sono di pari valore, evitare gli estremi di chi considera i mezzi a prescindere dai fini o i fini a prescindere dai mezzi. È dunque impossibile esprimere un algoritmo che permetta di prendere decisioni etiche: solo chi è già virtuoso capisce quello che deve fare. Il virtuoso non può dire al non-virtuoso come agire, può solo mostrargli la strada per diventare a sua volta virtuoso, cioè dargli il buon esempio che permetterà di sviluppare la *phronèsis*.

Ho espresso questa etica della virtù in uno stile diverso rispetto al resto del libro. Si tratta di una filosofia classicista, molto diversa da consequenzialismo e deontologia: rifiuta ciò che è comune alla discussione delle altre teorie e così risulta incommensurabile, non si può dire se sia una teoria più o meno adatta delle altre. Credo però che la virtù sia una componente reale dell'etica, irriducibile a doveri o conseguenze, ed è quindi necessario che anche le altre teorie ne rendano merito.

Ritengo che la deontologia kantiana prima esposta e questa etica della virtù possano formare un'unica teoria. Nello specifico, la virtù, all'interno di quella deontologia, ha tre ruoli. Primo, la virtù è il criterio aggiuntivo necessario a scegliere come agire nei casi di azioni non regolamentate (quelli in cui un'azione produce contesti accettabili sia se attuata da tutti, sia se non attuata da alcuno). Secondo, come conseguenza del primo ruolo, la virtù è anche il criterio della supererogazione, cioè le azioni supererogative, le azioni buone ma al di fuori del dovere, sono quelle virtuose. Terzo, gli scopi legittimi sono quelli virtuosi: cioè il realizzare l'*eudaimonia* è ciò che migliora la condizione del mondo e ciò che dà merito, la *phronèsis* permette di riconoscere ciò che lo fa e quindi di giudicare se i contesti di universalizzazione sono razionalmente accettabili, ne consegue che le motivazioni di primo livello che fanno da base alla buona volontà sono quelle *eudaimoniche* cioè quelle che riconoscono il valore intrinseco di ciò che si fa.

Personalmente, questo è ciò che io credo: l'azione buona è quella con uno scopo virtuoso tale da rispettare l'imperativo categorico.

LETTURE PER APPROFONDIRE

Parto con le letture rilevanti per il consequenzialismo. Tra i classici, i principi di tale filosofia si trovano già in Machiavelli (1532). Tra i contemporanei, si vedano Singer 1995, e Onfray 2007 e 2006-2018: quest'ultima è la sua contro storia della filosofia, un progetto in continua crescita sui pensatori non allineati di varie epoche.

Passo alla deontologia. Di Kant aggiungo 2004; personalmente, tra i classici, almeno per quel che riguarda l'idea di una legge naturale, vedo influenze deontologiche anche in Hobbes 1651, Vico 1725, Montesquieu 1748, Beccaria 1764. Tra i continentali, Habermas 1993. Per una visione conservatrice invece, Scruton 2007 e 2015. La più importante eticista puramente kantiana vivente è Korsgaard (1996).

Per l'etica della virtù, tra i classici bisogna recuperare i virili stoici, dai quali sono profondamente influenzato: ad esempio, Cicerone 2019, Marco Aurelio 2010, Epitteto 1996. Tra i moderni che continuano questa tradizione (ormai memeficata dalla rete), ha acquisito una certa popolarità il biologo italiano Pigliucci (2017). Tra gli psicologi che continuano a proporre una visione stoica, si vedano Seligman 2007 e Peterson 2018. Tornando nel campo dell'etica, le opere di Nussbaum (1997, 2011) mostrano influenze aristoteliche; si veda anche la sua collaborazione con il Nobel per l'economia Sen (Nussbaum & Sen 1993), il quale aveva già affrontato tematiche etiche con Williams (Williams & Sen 1982). Un altro premio Nobel per l'economia (ma psicologo di professione) rilevante per i temi dell'etica della virtù è Kahneman (2010).

Molti pensatori religiosi moderni sono a cavallo tra deontologia e virtù: ecco Hauerwas (1974, 2020), Adams (1997), Bochenski (Miche 1946), Zagzebski (1996, 2001, 2003). Tra tutti però, quello che dovrebbe attrarre di più l'attenzione è Matthieu Ricard (2003), genetista, figlio del filosofo Revel, monaco buddista, definito sulla base di esami neurologici «l'uomo più felice del mondo»: chi meglio di lui, allora, per parlare della virtù?

Etica: metaetica

Quarto ragionamento: i valori come base della conoscenza

In questo capitolo il ragionamento si concentra soprattutto sulla divisione tra fatti e valori; nello specifico, questo è lo schema:

- P1** affermare valori non esprime emozioni né ordini, è un tentativo di riferirsi a fatti;
- C1** quindi il non-cognitivismo è falso;
- P2** alcune affermazioni contenenti concetti spessi sono vere;
- P3** la parte valutativa delle virtù epistemiche (che sono spesse) non può essere negativa;
- P4** anche ruoli, funzioni e promesse sono fatti che implicano valori;
- P5** e alcune affermazioni contenenti ruoli e funzioni sono vere;
- C2** quindi il realismo deve essere vero;
- P6** non ci sono prove definitive a favore né del naturalismo né del non-naturalismo...
- P7** ...ma i fatti vengono individuati a partire dai valori;
- P8** pluralismo e soggettivismo non sono alternative a naturalismo o non-naturalismo (ma il relativismo è inconsistente);
- C3** quindi la scelta tra naturalismo e non-naturalismo dipende dai valori stessi.

Le varie tesi 1

La metaetica incrocia l'etica con le altre discipline filosofiche: si interroga sia su domande epistemologiche (come sia possibile acquisire conoscenza delle verità etiche) sia su domande metafisiche (quale sia la natura del bene). Quelle esposte di seguito sono tutte teorie descrittive, mai normative.

È comune distinguere tra *ontologia metaetica* e *semantica metaetica*, cioè tra domande sulla natura del bene (se sia oggettivo o relativo, se sia un valore o inglobi più valori) e domande sul significato del termine «bene» (se costituisca un'affermazione su fatti o meno, e a quali fatti cerchi di riferirsi). Per la prospettiva della filosofia analitica, si tratta di una divisione artificiale: io non posso dire, semanticamente, quand'è che una frase è vera senza avere, ontologicamente, una teoria sui fatti ai

quali la frase cerca di riferirsi, c'è unione tra ciò che il bene è e ciò che la parola «bene» significa. Per questo motivo preferisco presentare un'unica tassonomia delle teorie metaetiche:

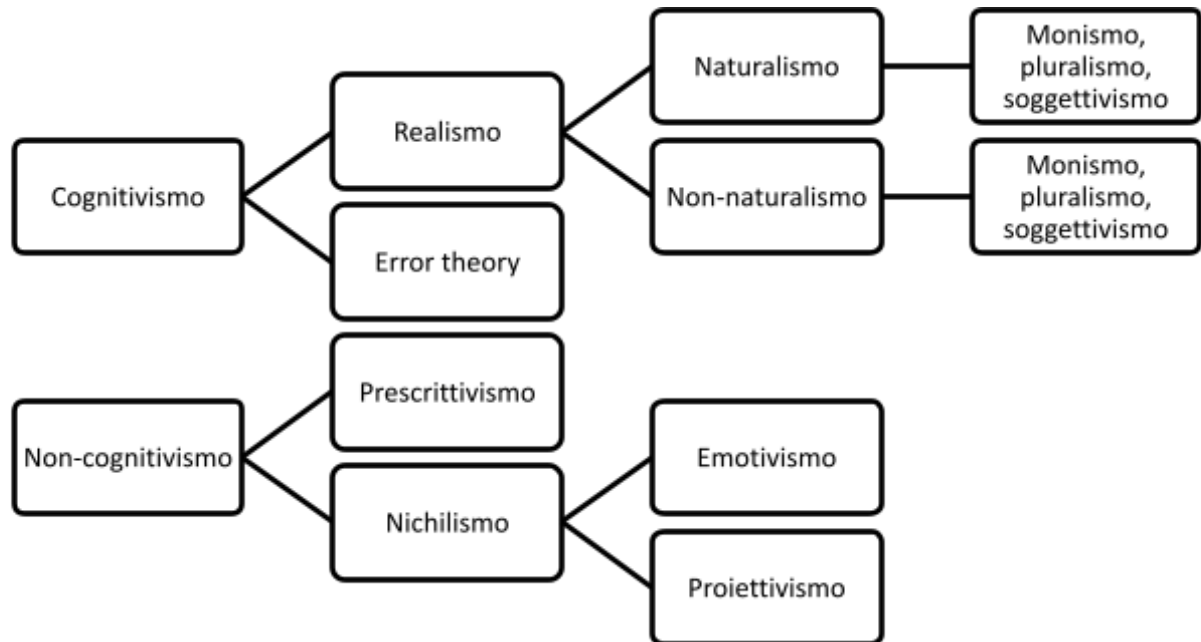


Figura 10

1. Le espressioni etiche sono proposizioni, cioè affermazioni su fatti, cioè frasi che possono essere o vere o false?
 - **Si:** sei un *cognitivista*, vai alla terza domanda.
 - **No:** sei un *non cognitivista*, vai alla seconda domanda.
2. Ci sono valori, norme etiche?
 - **Si:** sei un *prescrittivist* (n.b.: anche tutti i cognitivisti risponderebbero così).
 - **No:** sei un *nichilista*, vai alla quarta domanda.
3. Ci sono proposizioni etiche vere, cioè esistono fatti etici?
 - **Si:** sei un *realista*, vai alla quinta domanda.
 - **No:** sei un sostenitore dell'*error theory*.
4. Cosa fanno le espressioni etiche: esprimono direttamente emozioni (senza voler fare affermazioni), oppure parlano di emozioni *come se* fossero fatti, simulando, fingendo?
 - **Esprimono emozioni:** sei un *emotivista*.
 - **Simulano affermazioni:** sei un *proiettivista* (o *funzionalista* che dir si voglia).
5. I fatti etici sono naturali (e.g. strutture cerebrali), o sono fatti *sui generis*, di altra categoria?
 - **Fatti naturali:** sei un *naturalista*, vai alla sesta domanda.
 - **Fatti non-naturali:** sei un *non-naturalista*, vai alla sesta domanda.

6. Quanti valori non-contadditori ci sono? Questi valori sono dipendenti da condizioni personali, sociali, psicologiche, quali cultura, religione, sessualità, cioè ogni uomo va giudicato in modo diverso e indipendente?
- **Un valore, insensibile alle condizioni:** sei un *monista*.
 - **Più valori, sensibili o meno che siano:** sei un *pluralista*.
 - **Uno o più valori, sensibili alle condizioni:** sei un *soggettivista*.

Affermare valori non esprime emozioni né ordini: contro il non-cognitivismo

Il linguaggio racchiude molto di più delle sole affermazioni: «come stai?», «buongiorno», «vi dichiaro marito e moglie», «ti auguro buona fortuna», «ti offro un milione», «ti prometto fedeltà», «ti ordino di chiudere la finestra», «ti prego di tacere», «*nuntio vobis gaudium magnum*», sono tutte espressioni che non esprimono affermazioni, cioè che non cercano di riferirsi a fatti, bensì fanno altro, sono modi di compiere altre azioni con le parole (Austin 1975). Per i *non-cognitivist* le espressioni etiche sono di questo tipo: non sono né vere né false (né conoscibili quindi) perché non cercano di riferirsi a fatti.

A. J. Ayer è un peso massimo della filosofia del Novecento: una volta, ad una festa, Mike Tyson stava importunando Naomi Campbell, Ayer si è messo di mezzo, Tyson gli ha detto «sai chi sono? Sono il campione del mondo dei pesi massimi!», e lui «invece io sono il *Wykeham Professor of Logic*, siamo entrambi ai vertici nel nostro campo, dovremmo parlarne come uomini razionali», al che Ayer e Tyson si sono seduti a discutere di semantiche formali, mentre Naomi Campbell se l'è svignata (Rogers 1999). Ayer è il campione dell'*emotivismo*: secondo lui espressioni etiche come «uccidere è male» sono solo manifestazioni di emozioni, andrebbero tradotte con qualcosa come «abbasso l'uccidere, viva il salvare vite!» (Ayer 1952) Ciò è ben diverso dal dire «sento che uccidere è male, mi causa dolore»: questa espressione fa un'affermazione su *fatti emotivi*, io oggettivamente sento o non sento dolore; invece per l'emotivismo le espressioni etiche sono *atti emotivi*, un po' come urlare «urrà!»

Secondo Hare, velocissimo filosofo dalle lunghe orecchie, le espressioni etiche sarebbero invece *ordini universali*, «uccidere è male» starebbe a dire «che nessuno uccida!» Questo è il modo in cui l'imperativo categorico esprime i doveri, come ordini e non come affermazioni su fatti: Hare mantiene così la teoria vicina alla deontologia kantiana in quanto secondo lui l'universalizzazione è essenziale all'etica. In sintesi, questo *prescrittivismo* consente di preservare l'idea che le regole etiche siano uguali per tutti pur negando la loro fattualità (Hare 1952).

Queste teorie sono difese in modo pragmatico: se non ci sono fatti etici, il non-cognitivismo è l'unica alternativa all'*error theory*, e dire che un'affermazione etica non fa affermazioni su fatti è più utile e desiderabile del dire direttamente che è falsa. Ma tali teorie fanno a loro volta affermazioni su fatti linguistici: il problema, quindi, non è se vi siano o meno fatti etici, ma se davvero il parlante che dice «uccidere è male» intenda solo esprimere un'emozione o un ordine, e non fare anche un'affermazione su fatti.

Peter Geach (1965) ha formulato una potente obiezione contro il non-cognitivismo. Considero questi casi:

1. atteggiamenti proposizionali: «credo che uccidere sia sbagliato»;
2. congiunzioni: «l'erba è verde e uccidere è male»;
3. ragionamenti: «uccidere un essere vivente è male, un cane è un essere vivente, quindi uccidere un cane è male».

Una credenza è un atteggiamento proposizionale (Schwitzgebel 2019): se è così, nel primo esempio verso quale proposizione dovrei avere l'atteggiamento di credenza se non «uccidere è male»? Negare che le credenze siano atteggiamenti proposizionali per immaginare atteggiamenti tenuti verso ordini o espressioni d'emozioni peggiora la situazione: dire «credo: che nessuno uccida!» o «credo che abbasso l'omicidio!» sono chiare assurdità, e lo sono perché il contesto richiede una proposizione e non un ordine o un'emozione. Similmente il secondo esempio: «l'erba è verde» è senza dubbio un'affermazione, quindi anche il suo congiunto deve esserlo perché «l'erba è verde ed abbasso l'uccidere!» sarebbe assurdo. Stessa cosa il terzo esempio: da due proposizioni posso dedurre una terza, ma da un ordine o un'espressione d'emozione non posso dedurre alcunché, eppure il ragionamento sembra legittimo, quindi le espressioni etiche che lo compongono devono essere affermazioni.

Se si accetta tale argomentazione il non-cognitismo non è più un'alternativa valida all'error theory, e chi vuole evitarla deve passare al realismo. Oppure è un antirealista, e allora non si preoccupa: per gli antirealisti è sufficiente sposare una teoria deflazionaria della verità, così una espressione etica può essere usata in contesti che presuppongono valori di verità anche se non si riferisce a fatti (Horwich 1993). L'antirealismo è però una posizione con i suoi problemi, mantenere il realismo è desiderabile: la posizione ideale sarebbe quella in grado di accettare insieme il corrispondentismo, l'inesistenza di fatti etici, e il fatto che espressioni etiche hanno comunque valori di verità positivi.

Questo è ciò che cerca di fare la più importante variante del *proiettivismo*, il *quasi-realismo* sviluppato da Simon Blackburn, uno dei maggiori eticisti viventi. «Questo è un tavolo» esprime una proposizione, ed è un'affermazione su fatti, ma io posso guardarla con fare interrogativo, «questo è un tavolo?», e allora non è più un'affermazione. L'idea è che espressioni come «uccidere è male» siano quasi come domande, sembrano affermazioni ma in realtà si ha un diverso e particolare modo d'intenderle. Nello specifico, sono davvero manifestazioni d'emozioni come dice Ayer, ma queste emozioni sono *proiettate* sul mondo: quando dico che un film è *pauroso* parlo come se la paura fosse una caratteristica intrinseca del film, un fatto esterno, quando ciò che rende vera l'affermazione è invece la mia emozione; similmente, parlo del bene e del male come fossero caratteristiche del mondo, anche se in realtà sto esprimendo emozioni (pur causate da fatti del mondo).

Quindi le espressioni etiche, pur non essendo affermazioni, vengono usate *come se* lo fossero; ne consegue che, anche se non si riferiscono a fatti, è legittimo trattarle *come se* lo facessero, e quindi possono avere lo stesso un valore di verità. I fatti etici sono come i personaggi di un libro, sono dei miti o dei simboli: «Sherlock Holmes è un uomo» non cerca di riferirsi a fatti, ma è trattato come se lo facesse, ed in virtù di questa *utile finzione* è legittimo dire che è vero; stessa cosa i giudizi etici, li esprimo come fossero fatti solo per una strategia retorica volta ad essere più convincente, ma possono comunque essere detti veri (Blackburn 1993, 1998).

Detto in altro modo, la tesi del *quasi-realismo* è che l'atteggiamento proposizionale dei parlanti nei confronti delle affermazioni etiche sia sì antirealista, ma tale da mimare i normali rapporti logici. Per esempio, posso dire che «uccidere è male» e «uccidere non è male» sono contraddittorie non perché affermino fatti contrari, ma solo perché, se ho un atteggiamento di disapprovazione verso

Uccidere, non posso contemporaneamente avere nei suoi confronti un atteggiamento di approvazione. Dunque, la risposta a Geach è che posso comunque usare espressioni etiche in congiunzioni e ragionamenti perché il comportamento logico dipende dall'atteggiamento e non dai valori di verità.

Di conseguenza il quasi-realismo è *conservativo*: il negare l'esistenza dei fatti etici non richiede di cambiare il modo in cui si parla del bene, posso legittimamente continuare a parlarne come esistesse. Ciò però chiede dei sacrifici alla filosofia analitica: essendoci uno scarto *sistematico* tra ciò che uno dice e ciò che uno intende, diventa quasi impossibile capire la mente tramite il linguaggio, un obiettivo che sta al cuore di questa tradizione e che caratterizza tutto il suo metodo di lavoro.

Ho alcune obiezioni. Anzitutto, che si tratti di una teoria quasi-*realista* è già dubbio. Se non ci sono fatti etici, e la verità sta nella fattualità, non dovrei dire che le espressioni etiche sono vere, dovrei dire invece «sono false, o né vere né false, anche se ne parlo come fossero fattuali». Se un parlante realista crede nell'esistenza di fatti etici, allora non parla *come se* vi fossero, cerca di riferirsi a quei fatti; e se uno non vi crede, e parla *come se* vi fossero, sa di dire falsità, non verità, sta solo *usando* la falsità per guidare il comportamento. Quando io dico che il film è pauroso, fallisco nel riferirmi a delle caratteristiche del film, ma ci sono comunque altri fatti, cioè le mie emozioni, che rendono l'affermazione vera (per di più, la mia emozione è causata davvero da caratteristiche oggettive del film, caratteristiche che se non vi fossero state il film non sarebbe stato pauroso); ma se non ci sono fatti etici di alcun tipo (neanche naturali come le emozioni), non c'è altro che costituisca la verità di «uccidere è male». Quindi sembra che non vi sia alcun motivo per il quale certe frasi debbano essere riconosciute come vere se non in una interpretazione puramente antirealista.

Poi, anche accettando la tesi del quasi-realismo, questa non riuscirebbe a superare l'argomentazione di Geach nel contesto della negazione. Prendo due credenze: «uccidere è male» e «uccidere non è male». Poiché il quasi-realista dice di avere nei confronti della proposizione un atteggiamento di disapprovazione (e non una credenza fattuale), la prima è da intendere come «disapprovo l'uccidere». Ma quindi, qual è la sua negazione? Sarà o «non disapprovo l'uccidere» o «disapprovo il non-uccidere», ma nessuno di questi due atteggiamenti è l'equivalente dal credere che «uccidere non è male» («non disapprovo l'uccidere» è equivalente al non-credere «uccidere è male», «disapprovo il non-uccidere» è equivalente al credere «non-uccidere è male»). Quindi il problema sollevato da Geach rimane intatto (Unwin 1999, Dreier 2006, Schroeder 2008).

Una difesa influente di Gibbard (1990, 2003) sta nell'interpretare il contenuto delle proposizioni etiche come insiemi formati da mondi possibili da una parte e regole di comportamento dall'altra: quando dico «uccidere è male» escludo tutte le coppie che prevedono regole che non vietano l'uccidere, quando dico «uccidere non è male» tengo solo quelle che ho escluso prima, e quando dico «uccidere è male e uccidere non è male» mi contraddico perché l'intersezione dei due insiemi è vuota. Questo riesce effettivamente a spiegare, in modo molto originale e accettabile dal quasi-realismo, come interpretare la negazione e assegnare un valore di verità a congiunzioni di espressioni etiche, ma non è affatto chiaro come si dovrebbero gestire le congiunzioni di espressioni etiche e non-etiche.

Più in generale, ci sono vari motivi per sostenere che i giudizi etici non sono basati solo sulle emozioni, come vogliono emotivismo e quasi-realismo. Il disgusto che provo davanti ad un'azione è un'emozione negativa, eppure non tutto ciò che è disgustoso (e, in generale, emotivamente spiacevole) lo riconosco come male, deve esservi dell'altro. Se chi parla cerca solo di offrire un commento emotivo, e non anche di fare affermazioni fattuali sul mondo esterno, perché i parlanti trattano chi ha opinioni etiche diverse dalle proprie come se dicesse che il cielo è verde? Perché

davanti a chi dice «stuprare bambini è bello» si risponde con la denuncia, e non dicendo «vabbè, i gusti son gusti» o «io ho le mie emozioni e tu hai le tue»? E se cambio idea, perché dico «mi sono sbagliato, hai ragione, è bello», perché dico di aver errato, e non solo «le mie emozioni sono cambiate», in modo neutrale? Forse allora sto esprimendo ordini? Se io mi chiudo nella mia cameretta e penso «Caligola ha fatto male a far uccidere i suoi servitori» a chi starei dando ordini, di chi starei cercando di influenzare il comportamento, visto che nessuno può sentirmi? (cfr. Brandt 1959) Tutti questi esempi non lasciano dubbi: il parlante cerca davvero di riferirsi a fatti etici.

Queste teorie individuano correttamente un elemento del giudizio etico: quando emetto un giudizio come «uccidere è male» esprimo anche un'emozione (e dò anche un ordine). Questo giudizio, inoltre, è davvero influenzato dalle emozioni che provo davanti all'azione giudicata. Non posso dire, però, che quel giudizio non sia nulla oltre all'emozione. Davanti ad un quadro di Tiziano, provando una certa emozione, posso esclamare «bello!» Poi, però, mi fermo, cerco le linee di costruzione, rifletto sulla struttura, sulla tecnica, sul motivo per cui i personaggi sono rappresentati in quelle pose anziché altre, e così alla fine potrei dire «mi piace ma non è bello». L'emozione immediata è distinta dal valore, l'emozione che dà un omicidio può farmi urlare «è male!» ma non è questo che *lo rende* male, quell'esclamazione non è ancora il giudizio, mi stimola e mi dà una direzione ma devo ancora elaborare razionalmente (elaborazione razionale che causerà a sua volta nuove emozioni).

Emozione e ragione hanno ruoli di pari importanza nel giudizio etico. La psicologia conferma che emozione e ragione non sono due processi opposti, né sono divisi in modo netto: senza emozione non vi sarebbe neanche ragione (Damasio 1995), non a caso le persone più intelligenti e razionali sono anche quelle più emotive e amichevoli (Mayer & al. 2004). È vero che molti non ragionano sulle proprie emozioni etiche, ed è vero anche che il mio cervello si è evoluto in modo tale che io sono propenso a considerare male l'omicidio a prescindere dalla verità di tale idea (Brosnan 2011); ma questo è irrilevante perché io ho comunque la possibilità di ragionare sulle mie intuizioni in modo dipendente dalla loro verità.

Nel secolo scorso il non-cognitivismo era una posizione molto comune grazie all'influenza dei positivisti: per loro tutte le espressioni etiche, essendo empiricamente inverificabili, erano insensate e quindi incapaci di riferirsi a fatti. Oggi il non-cognitivismo è una posizione fortemente minoritaria (Bourget & Chalmers 2014).

Le varie tesi 2: fatti e valori

Se le espressioni etiche sono affermazioni, ma non esistono fatti etici, allora bisogna accettare l'error theory: esiste questo enorme campo del discorso che è l'etica, ma è composto tutto da proposizioni false, come il discorso dell'astrologia.

L'error theory non spiega perché, se una proposizione etica è falsa, si dovrebbe adottarla come credenza invece di rifiutarla; epperò non lo impedisce neanche: un naturalista potrebbe affermare che « $2+2=4$ » è fattualmente falso (perché ci sono fatti naturali, non fatti matematici) ma può comunque essere utilizzato come aiuto per il ragionamento e la modellazione dei sistemi fisici; similmente, «uccidere è male» è fattualmente falso ma può essere utilizzato per coordinare il comportamento. Il fatto che una norma sia falsa non significa che io possa disubbidirvi (Joyce 2016).

Uccidere è male, ho detto? Perché? Perché è la mia natura, nessuno pensa che sia bene. Ma arriva lo scettico che risponde: «è vero, pensiamo così, ma non è quello che *dovremmo* pensare». Allora lo è perché le leggi puniscono l'omicidio; e lui: «non *dovrebbero* punirlo». È male perché chi uccide viene punito; «non *dovremmo* punirlo». Ma finché ci sono queste leggi, poiché nessuno vuole essere punito, nessuno deve uccidere; «non *dovremmo* non voler essere puniti». È male perché ha conseguenze dolorose; «non *dovremmo* percepirlo come doloroso, ha conseguenze negative ma ciò non significa che non vada attuato». È male perché Dio ha comandato di non uccidere; «l'ha comandato, ma noi non *dovremmo* fare ciò che dice, oppure Lui *avrebbe dovuto* dire diversamente».

Perché uccidere è male? Io sto portando a supporto dei fatti, dei modi in cui il mondo è, dei modi in cui le cose *esistono*. Il bene e il male, però, non descrivono le cose *così come sono*, bensì indicano le cose *così come dovrebbero essere*: a qualsiasi fatto io possa portare sarà sempre possibile rispondere «questo è il mondo com'è, non come deve essere». Nota Hume (1739) che, posto un qualsiasi insieme di proposizioni della forma «x è y» o «*esiste* x», non potrò mai inferire logicamente una conclusione della forma «x *dovrebbe essere* y», «*dovrebbe esserci* x», cioè non esistono implicazioni del tipo «se x è y, allora *dovrei* fare z». Dal fatto che il mondo sia in un certo modo non si può mai derivare un dovere.

In altre parole, posso dimostrare l'*error theory* così: la *distinzione tra l'essere e il dover essere* implica la *distinzione tra fatti e valori*, quindi immaginare l'esistenza di *fatti etici* è un'insensatezza, cioè nessuna affermazione etica può essere fattualmente vera. Questo è il problema fondamentale di tutta la metaetica.

Il ragionamento di Hume poggia sul rifiuto di una intuizione di Platone. La geometria descrive triangoli perfetti, diversi da tutti i triangoli grezzi e storti che esistono in natura: come sono arrivato a dedurre questa idea del triangolo, il triangolo *così come deve essere*? La perfezione è una generalizzazione: tutti i quasi-triangoli naturali sono categorizzati in base a caratteristiche comuni, caratteristiche che mi permettono di riconoscere un triangolo quando lo vedo; queste caratteristiche, isolate, creano il triangolo perfetto. E questo non vale solo per la geometria: per esempio, anche gli alberi li classifico e identifico in virtù di certe caratteristiche comuni, quindi esiste una *idea di albero* che costituisce l'*albero perfetto*. E se penso al mondo tutto, ecco l'*idea del bene*: sperando il mondo *così com'è*, imperfetto, inferisco il mondo *così come dovrebbe essere*, e quindi i valori e i doveri che puntano ad esso. Comprendere come dovrebbe essere il mondo è intuitivo come guardare il quadro di un artista dilettante e capire subito quali sono i suoi e cosa avrebbe dovuto fare per evitarli. In Platone, quindi, non c'è divisione tra fatti e valori, tra il vero e il bene (Platone 2009 Timeo): che da un «è» sia possibile dedurre un «devo» è manifesto nell'esperienza, ed è così perché, dato che il mondo come appare è illusorio e solo il mondo delle idee perfette esiste in senso pieno, *ciò che è* e *ciò che deve essere* coincidono.

Chi ha ragione: Hume o Platone?

Alcune affermazioni contenenti concetti spessi sono vere 1: argomentazioni a favore dell'indivisibilità

Penso ai concetti della virtù, penso a parole come «coraggioso», «temperante», «magnanimo». Questi termini esprimono *concetti spessi*, e ciò li distingue da «bene» che esprime un *concetto fino*. I concetti spessi sembrano specificazioni di quello fino, cioè «coraggioso» starebbe a dire «bene *in un certo modo*», lo è ciò che ha certe caratteristiche che sono anche bene. Quindi, mentre il concetto

fino esprime solo valore, il concetto spesso esprime sia fatti che valori, dice sia come si è agito sia che è stato bene, ha una componente *descrittiva* e una *valutativa*.

Alcuni filosofi credono che anche i concetti fini debbano avere una parte descrittiva e una valutativa (o un'unica componente che include entrambi gli aspetti) (Hare 1963). Questo è motivato dalla speranza dei non-cognitivisti di poter superare l'obiezione di Geach: se anche i termini fini avessero due componenti, le relazioni logiche sarebbero giustificate dalla sola componente descrittiva. È però indesiderabile sia così: in tal caso, infatti, l'aspetto valutativo risulterebbe logicamente equivalente ai fatti indicati dalla parte descrittiva di «bene» (o «male»), e questo potrebbe essere giustificato dal fatto che il valore è riducibile a quei fatti, facendo risultare il non-cognitivismo falso.

Ora, se le due componenti sono distinguibili, se posso dire «coraggioso» intendendo solo la sua parte descrittiva a prescindere dalla valutazione, allora fatti e valori rimangono separati; ma se non è così, non è più legittimo sostenere una distinzione tra fatti e valori, vuol dire che certi fatti sono sempre e solo concepibili come implicanti un certo valore, e quindi c'è un valore fattuale. Dunque valutazione e descrizione sono unite?

McDowell, il più importante filosofo sudafricano vivente e raro *hegeliano analitico*, risponde di sì. Se la componente descrittiva e quella valutativa del concetto spesso fossero separabili, se a certi fatti non fosse associato per forza un certo valore, io dovrei essere in grado di individuare l'estensione del termine (la serie di atti per i quali è vero) anche conoscendo solo la parte descrittiva. Epperò, non può essere così. Infatti la parte descrittiva e quella valutativa del termine spesso hanno diverse estensioni: un atto coraggioso potrebbe essere quello che «è fatto prendendosi dei rischi, ed è bene», ma non tutti gli atti fatti prendendosi dei rischi sono bene, e non tutti gli atti che sono bene sono compiuti prendendosi dei rischi. Ne consegue che se io comunico ad una persona solo la parte descrittiva del termine «coraggioso», quella persona considererà come estensione del termine tutti gli atti fatti prendendosi rischi: ma questo è un fraintendimento, non è quello il senso della parola. Quindi, per essere competente nell'uso di un concetto spesso bisogna conoscere sia la parte descrittiva che quella valutativa, quindi non sono separabili (McDowell 1998).

Questo, però, non è sufficiente ad affermare l'inseparabilità delle due componenti: se hanno una estensione diversa vuol dire che indicano due diverse proprietà, l'estensione del termine spesso è l'unione delle due, ma ciò suggerisce proprio il contrario di quanto si vorrebbe, suggerisce che, essendo legittimo immaginare un'azione fatta prendendosi rischi ma che è male, la parte descrittiva *non* implica la valutativa.

Bernard Williams, l'eticista fisicamente più simile al cantante Leonard Cohen, porta un'argomentazione diversa. Immagina una *società ipertradizionalista*, una società ideale *omogenea* e *irriflessiva*, cioè tale che tutti i suoi componenti credano le stesse cose e nessuno vi rifletta su per metterle in dubbio. Per Williams (1985) i concetti di bene e male sono generalizzazioni, astrazioni, quindi emergono solo tramite riflessione, o meglio la riflessione avviene tramite questi concetti fini: osservo certe azioni, individuo i *pattern* di caratteristiche condivise, e battezzo questa parte comune «bene». Ma la società ipertradizionalista non riflette, quindi non usa concetti fini come *bene* e forse non li può neppure possedere finché non inizia a riflettere. Usa però concetti più specifici, che non richiedono generalizzazione, cioè ragiona sull'etica usando *solo* concetti spessi. Ne consegue che in tale società è impossibile formare frasi che abbinino concetti spessi e fini, è impossibile dire «è coraggioso ma non è bene»; di contro, anche se un antropologo apprendesse la lingua parlata da tale civiltà, potrebbe non riuscire a trovare un equivalente nelle lingue della civiltà occidentale per tali concetti etici.

Immagino però che quella società inizi a riflettere fino a sviluppare i concetti fini. Ora può formulare una domanda come «essere coraggiosi è bene o male?» Ma secondo Williams il «coraggiosi» di questa frase non esprime lo stesso concetto che esprimeva prima dello sviluppo dei concetti fini: infatti, può essere vero che in tale società «coraggioso» indichi una caratteristica positiva, ma ciò non implica «quindi è bene» in quanto tale società non possiede tale concetto; così, dicendo «essere coraggiosi non è bene» creo un nuovo concetto, e non nego la verità di quanto creduto in quella società. Il *vero* concetto spesso, che *non* può voler dire «bene in un certo modo» (com'è invece col mio), poteva essere utilizzato solo prima dello sviluppo dei concetti fini; nel momento in cui li ho sviluppati ho anche perso il concetto spesso, e ho perso tutta la conoscenza che vi era associata. Come spesso accade, ci sono cose che sembrano perfettamente chiare solo finché non si inizia a riflettere, se ci si pensa diventano subito oscure.

Quindi, in una società ipertradizionalista, la parte descrittiva e la valutativa sono inseparabili e indubitabili in quanto non esistono i termini fini necessari per farlo; ma nel momento in cui acquisisco la capacità di separarle, non esiste neanche più il concetto, cambio il modo in cui valuto. Quindi è vero che dalla parte descrittiva non si può inferire la parte valutativa, ma non perché le due siano separabili, al contrario, è proprio perché i *veri* concetti spessi, per definizione, sono uniti, e posso inferire una parte dall'altra solo dopo averli distrutti.

Il problema principale di questa argomentazione sta nel presupporre il *centralismo dei concetti spessi*, cioè l'idea che i concetti fini siano derivati dagli spessi, e quindi che si possa comprendere un concetto fino solo dopo aver compreso gli spessi. Se i concetti fini sono generalizzazioni degli spessi, ciò è corretto; epperò pare improbabile perché non si capisce come sia possibile esplicitare il significato dei termini spessi senza poter già usare termini fini. Se il centralismo è falso, allora non c'è motivo per il quale la capacità di distinguere la parte descrittiva dalla valutativa debba distruggere il concetto spesso, e quindi non c'è motivo per negare la distinzione tra fatti e valori (cfr. Dancy 1995, Tappolet 2004, Väyrynen 2013).

Passo infine ad un'argomentazione presa da Putnam (1985). Io intendo il termine spesso «onesto» come «chi non mente e non ruba, ed è bene che non lo faccia» (a sinistra della virgola c'è la parte descrittiva, a destra la valutativa). La parte descrittiva ha un ruolo caratterizzante: se qualcuno usasse il termine mantenendo la parte valutativa ma cambiando la descrittiva (o anche cambiando tutte e due) starebbe senza dubbio usando una parola diversa. Immagino, allora, una società che mantiene la parte descrittiva ma cambia la valutativa, cioè una società che ritiene l'onestà un male. In questa società si potrebbero dire, senza alcuna ironia, «costui ha una deprecabile tendenza all'onestà». Un'affermazione come questa non è logicamente contraddittoria, ma è assurda, costituisce un uso troppo diverso da quello abituale: se, durante un viaggio, sentissi un indigeno pronunciarla, penserei, come nel caso dell'acqua* dell'uomo* della Terra*, che anche la parola «onestà» debba voler dire *onestà**, cioè debba essere un'altra parola, con un diverso significato, rispetto a quando la pronuncio io. Quindi l'aspetto valutativo, almeno in alcuni casi, non è modificabile a piacere, bensì è dipendente dall'aspetto descrittivo.

Questo legame tra descrizione e valutazione, però, è *pragmatico*, non *semantico*: l'uso della parola «onesto» in senso negativo è inaccettabile per via di una convenzione, e non per l'inseparabilità delle sue componenti. Io posso usare una descrizione a scopo valutativo (posso dire «quello lì ruba...» con tono di biasimo), oppure una valutazione con scopo descrittivo («è un buon uomo» intendendo che è amichevole, aperto, diligente, equilibrato...), ed è quindi vero che da una descrizione si può inferire una valutazione, ma si tratta di una inferenza contingente, ha motivazioni sociali (non logiche) tali che in una società diversa potrebbe benissimo non sorgere, perciò fatti e valori rimangono distinti (Väyrynen 2013).

Alcune affermazioni contenenti concetti spessi sono vere 2: evidenti casi di divisibilità

Nessuna delle argomentazioni proposte si è dimostrata in grado di sostenere l'indivisibilità di descrizione e valutazione. Dall'altro lato del fiume, gli esempi e le argomentazioni contro abbondano.

Gli autori finora citati, sia pro che contro, notano che certi concetti spessi hanno un'esistenza limitata a specifici sistemi di valori, e chi respinge quel sistema, anziché modificare il loro valore, si rifiuterà direttamente di utilizzarli. Per esempio, se devo valutare una commedia erotica, e la disapprovo, dirò che è «sporca»; ma se la approvo, dirò che è «esplicita», non mi metterò a cercare di rivalutare positivamente «sporca». Il problema è che «sporca» ed «esplicita» hanno una identica parte descrittiva e una diversa parte valutativa: è vero che da «sporca» inferisco «quindi è male» e da «esplicita» inferisco «quindi non è male», però, da una pura descrizione come «è una commedia piena di donne nude e giochi di parole a sfondo sessuale» non posso inferire né «è sporca» né «è esplicita», quindi la parte descrittiva non implica la valutativa.

Questo modo di usare i concetti spessi manifesta la natura pragmatica e non logica dell'implicazione tra descrizione e valutazione (Väyrynen 2021). L'affermazione esprime una proposizione che è o vera o falsa, la domanda invece non lo fa perché *blocca* i valori di verità associati al concetto. Per esempio, se io so che Maria è una donna, e voglio dire solo verità, rifiuterò l'uso del concetto *uomo* nell'affermazione «Maria è un uomo», ma posso comunque essere disposto ad usarlo nella domanda «è Maria un uomo?» Ora, io rifiuterò l'affermazione «Maria è scostumata» sia se penso che Maria sia invece casta, sia se ho una visione positiva del suo comportamento: riconosco che il concetto implica una valutazione negativa della libertà sessuale, il valutarla positivamente è sufficiente a rifiutare l'uso del termine. Il problema, però, è che, se la valuto così, rifiuterò di usare il termine «scostumata» in *tutti* i contesti, anche nella domanda «è Maria scostumata?» Perché rifiuto quel concetto, se la domanda blocca i valori di verità? Risposta: perché la componente valutativa è indipendente dai valori di verità, solo la parte descrittiva li determina; l'affermazione «Maria è scostumata» è vera in virtù di un fatto, un comportamento, che poi la libertà sessuale sia cosa positiva o negativa è irrilevante alla verità. La stessa cosa vale per il caso paradigmatico degli insulti: è vero che non posso dire «è uno stronzo ed è bene che lo sia», ma ciò dipende solo da una convenzione sull'uso del termine, non da necessità logica. In altre parole: il termine spesso è semanticamente solo descrittivo, e solo pragmaticamente anche valutativo.

Infine, ci sono concetti spessi, come *frugalità*, il cui valore cambia col contesto: una «casa frugale» è rustica e piacevole, un «albergo frugale» è spartano e poco accogliente. Di nuovo, ad una identica descrizione si abbinano due diversi valori. È chiaro che diverse persone hanno diversi gusti, possono usare una stessa parola con scopi diversi anche in uno stesso contesto (per alcuni una casa «minimalista» è bella, per altri è brutta), oppure possono formulare una parola nuova che esprima meglio il loro giudizio; non pare però esservi alcun errore logico né nell'usare «frugale» in modo positivo, né nell'usarlo in modo negativo, quindi la sua descrizione non può implicare alcuna valutazione.

La parte valutativa delle virtù epistemiche non può essere negativa

Quello che posso dire sulla scostumatezza, però, non posso estenderlo a tutti i concetti spessi. Frasi come «ho fatto bene ad essere scostumato» od «ho fatto male ad essere coraggioso» suonano

strane ma sono logicamente accettabili, una frase come «ho fatto male ad essere saggio» invece suona assurda e inaccettabile quasi tanto quanto «ho fatto male a fare il bene». Hare, come visto, ipotizza che tutti i concetti abbiano una parte descrittiva e una valutativa, ma non mette sempre queste due componenti alla pari: in «scostumato» la parte descrittiva è più importante della valutativa, nel caso di «saggio» è invece la valutativa a dominare, per questo «è male essere saggio» suona paradossale. Ma personalmente, anche se è possibile che tutti i concetti abbiano entrambi gli aspetti, credo che il problema sia un altro.

La *saggezza* è una *virtù epistemica*, quindi un concetto spesso che fa da ponte tra conoscenza ed etica. In quanto segue intendo la sua parte descrittiva come «credere il vero, credere solo ciò che è giustificato, ponderare le decisioni» più varie altre caratteristiche. Ora affermo: «essere saggi è male». È come se dicessi «credere il vero è male, non bisogna farlo, bisogna invece credere il falso». Ovviamente, se credo che sia male credere il vero, credo anche che ciò sia vero, e quindi che sia male crederlo, cioè credo di *dover credere* che sia bene credere il vero. In altre parole: o credo che sia bene essere saggi, o, se non lo credo, ammetto di essere immorale perché credo di doverlo credere. Al che, visto che non voglio essere immorale, devo convertirmi e iniziare a credere che essere saggi sia bene. Posso non farlo? Questo è un problema perché allora sono costretto a proseguire i miei ragionamenti così:

- è male credere il vero, ed è vero, quindi è male credere sia male credere il vero;
- quindi dovrei credere che sia bene credere il vero;
- ma è vero che dovrei credere che sia bene credere il vero, quindi è male;
- quindi dovrei credere di dover credere che sia male credere il vero, ed è vero, quindi male;
- quindi dovrei credere di dover credere di dover credere che sia bene credere il vero...

e via così all'infinito. In sunto: se io credo di non dover credere il vero, qualsiasi cosa crederò vera, dirò di doverla credere falsa; ma, credendo sia vero che devo crederla falsa, dirò di doverla credere vera, quindi falsa, quindi vera eccetera. Sono finito in un *loop* razionalmente inaccettabile: l'unica possibilità è accettare che l'essere saggi sia bene. Posso dire che tra la descrizione e la valutazione di *saggezza* non c'è implicazione logica, ma c'è qualcosa in più dell'implicazione pragmatica, c'è una sorta di implicazione razionale: fatti e valori sono logicamente divisibili, ma razionalmente uniti.

Argomentazioni di questo tipo, che tentano di dimostrare che se si rinuncia ai valori etici si rinuncia anche alla possibilità della conoscenza, vengono chiamate *companions in guilt*, e sono tra le più discusse degli ultimi anni (soprattutto da Cowie, e.g. 2018).

Nonostante l'argomentazione sia suggestiva, la conclusione pare frettolosa: il ragionamento dimostra l'impossibilità razionale del considerare la saggezza come un male, ma ciò non implica che sia irrazionale considerarla *non-bene*, potrebbe essere né bene né male. Personalmente, credo che immaginare dei *gap* nei valori di verità sia una cattiva mossa, preferisco sostenere che tutte le proposizioni hanno il loro valore di verità; però, sono molti i logici che credono diversamente. Quindi, per difendere l'unione tra descrizione e valutazione in *saggezza*, non basta dimostrare la falsità dell'idea che sia un male, bisogna dimostrare la verità del suo essere un bene, bisogna dimostrare che è paradossale non considerarla tale.

Ruoli, funzioni e promesse sono fatti che implicano valori, e alcune affermazioni che li contengono sono vere

Io dico «prometto di portarti il libro», e l'amico pensa «ha promesso, quindi deve farlo». La promessa è un fatto, è il mio pronunciare e pensare certe cose, e, per definizione, è anche qualcosa che io *devo* fare: «non dovresti fare ciò che hai promesso» sarebbe come dire «non dovresti fare ciò che dovresti fare». La promessa, quindi, è un fatto che implica automaticamente un dovere cioè un valore deontologico (Searle 1964).

Passo ora ai ruoli, o alle *istituzioni* come dice Searle (2006), che sono definiti da doveri, e quindi implicano valori. Per fare un esempio ispirato a Prior, dal fatto che «Schettino è il capitano della nave» inferisco «Schettino ha il dovere di salire sulla scialuppa di salvataggio per ultimo»: questo dovere è intrinseco nella definizione di capitano, dire che il capitano non dovrebbe fare così è autocontraddittorio perché se non deve non è il capitano (MacIntyre 2007).

Ci sono poi le funzioni. MacIntyre (2007) fa l'esempio dell'orologio: dal fatto che «questo orologio perde minuti» si inferisce il giudizio «questo non è un buon orologio». L'orologio è definito dalla sua funzione, dal suo indicare l'ora, quindi obiettare «è corretto che perda minuti» sarebbe assurdo: un oggetto pensato per perdere minuti non è un orologio nel senso normalmente inteso.

Altri esempi di questo tipo abbondano in Putnam e Wittgenstein, ma una volta tradotti in italiano molti non suonano più bene. Wittgenstein (1967) fa partire un suo ragionamento etico dalla formula «una buona sedia»: una sedia è buona se è *funzionale*, è buona se mi regge e non collassa, cioè se fa ciò per cui è stata creata; generalizzando, il bene è relativo agli scopi, un'azione è bene se avvicina allo scopo, un bene assoluto è inesprimibile e incomprensibile come un miracolo. Ma in italiano la *bontà funzionale* non è la *bontà morale*: la frase «un buon orologio» è accettabile perché interpretabile nel senso funzionale «un orologio che funziona come previsto», invece «un orologio buono» è assurdo perché richiede un'interpretazione morale come «un orologio di gran cuore». Bisogna considerare, infatti, che in inglese (come in tedesco) non c'è distinzione tra il *bene* e il *buono*, c'è solo il «good»; così, concezioni facilmente formulabili in italiano, come «il buono non è il bene» o «un buon orologio non è un orologio buono», sono invece difficili da rendere in inglese; così, una frase come «a good chair», che in inglese pare moralmente rilevante, in italiano non lo è. Questo mostra che la filosofia analitica può avere varianti regionali: le argomentazioni che funzionano in una lingua non sono sempre riproponibili in un'altra.

Noto allora che in italiano il termine «buono» mostra un comportamento particolare se applicato all'uomo: quando usato per qualificare oggetti, se precede il nome («un buon tavolo») ha significato figurato funzionale accettabile, mentre se lo segue («un tavolo buono») ha significato morale inaccettabile; però quando lo si usa per qualificare una persona in entrambi i casi («un buon uomo», «un uomo buono») ha solo significato morale, mai funzionale. Mentre nella filosofia antica l'uomo aveva una *natura*, uno scopo ovvio come può averlo un martello, oggi, alla luce della biologia moderna, è difficile immaginare una funzione che possa definire l'umano: se voglio trasformare una sedia malmessa in «una buona sedia» so cosa devo fare perché conosco la sua funzione, ma se mi chiedo come potrei trasformarmi in «un buon uomo» non so che pesci pigliare. Se l'uomo avesse una funzione, uno scopo naturale e assoluto indipendente dal suo ruolo e dalla sua volontà (perché il fatto che io mi ponga lo scopo di diventare ladro deve essere irrilevante), allora il realizzare tale funzione sarebbe un bene di fatto; ma non è detto la abbia.

Comunque, promesse, ruoli, e funzioni (applicabili all'uomo o meno che siano), implicano valori i quali, entrando a sistema con le altre credenze, danno ad altri fatti la facoltà di implicare ulteriori valori. Per esempio, ipotizzo di aver derivato dalle funzioni il principio «causare dolore è male».

Ora, io non so se un'aragosta gettata viva nell'acqua bollente provi dolore, ma se è così, il suo sentire dolore è un fatto, un fatto che entrando a sistema col principio detto implica che è male bollire l'aragosta viva. Quindi l'idea che i fatti siano *moralmente irrilevanti*, cioè che i fatti non debbano influenzare la formazione delle credenze morali (come deve essere se fatti e valori sono separati), è un'idea insensata, al contrario sono i principi morali stessi che, essendo relazioni che associano giudizi a fatti, li richiedono (cfr. Stratton-Lake 2020).

Per questo, un irriverente Putnam (2004) arriva ad Hume di autocontraddirsi: Hume stesso, dalla distinzione tra fatti e valori (che ritiene essere un fatto), deduce come minimo che si deve evitare lo sforzo di dedurre valori da fatti, e questo è un valore; per di più, il ragionamento che lo ha portato a concludere la distinzione tra fatti e valori è possibile solo tramite una logica, la quale per i neopragmatisti come Putnam è normativa, e quindi presuppone a sua volta valori.

Come il noncognitivismo, anche l'error theory etica ha oggi pochi sostenitori (Bourget & Chalmers 2014): è evidente che almeno alcune affermazioni etiche devono essere vere, ritenere falsa una frase come «il capitano deve scendere dalla nave per ultimo» significa non capire cosa voglia dire l'essere capitano.

Detto tra parentesi: difesa dell'argomentazione di Searle

Riassumendo:

- le promesse implicano doveri;
- certi ruoli (come quello di capitano) implicano doveri;
- l'essere definiti da una funzione (come quella di segnare l'ora) implica doveri;
- alcune virtù (soprattutto epistemiche) descrivono caratteristiche che non possono essere giudicate negativamente.

Quindi, se io dico «devo fare questo» e lo scettico mi chiede «perché?» posso rispondere «l'ho promesso, sono un capitano, è il naturale scopo di un essere umano, è virtuoso», e se risponde «questo non è bene» posso ribattere «e il tuo credere così ti rende immorale».

Ma ecco sollevarsi Charles Pigden (2016), che porta un'obiezione potente e utile contro tale dottrina. Il suo bersaglio principe è Searle il quale, difendendo l'idea che le promesse implicano doveri, sostiene:

- se *compio l'atto linguistico del permettere* su x , devo fare x .

Cosa vuol dire compiere un atto linguistico? Pigden lo analizza così:

- se dico «prometto x », *secondo il gioco del promettere*, devo fare x .

Stare alle regole del gioco linguistico del promettere vuol dire accettare che se prometto qualcosa allora devo farlo, cioè Pigden sostiene che l'implicazione di Searle necessita di una seconda premessa per essere valida:

- se prometto x , *e devo fare quanto prometto*, allora devo fare x .

Se questa interpretazione è corretta, l'esempio di Searle non dimostra l'unità di fatti e valori in quanto, tra le premesse, ci sono sia constatazioni di fatti che affermazioni etiche. Cioè io, pur comprendendo perfettamente il significato di «promettere» e del gioco linguistico associato, posso credere che tale gioco linguistico sia falso o immorale: quindi, pur capendo quanto detto, non crederei di dover fare quanto promesso, e di conseguenza tra promessa e obbligo non può esservi analiticità. Ma è davvero così, e davvero l'atto linguistico è solo il pronunciare una certa frase all'interno di un certo gioco cioè di una certa convenzione?

Prendo come esempio il matrimonio. Quand'è che io sono sposato? Direi, come requisito minimo, quando un certo numero di persone-chiave (in particolare chi produce i miei documenti) mi riconosce come marito di *x*. In questo senso, il matrimonio è un fatto oggettivo, è oggettivo se il sindaco mi pensi come sposato oppure no. Allora dirò che le parole «vi dichiaro marito e moglie» costituiscono un atto di dichiarazione ogni volta che a pronunciarle è un funzionario tale che questo suo pronunciarle sia sufficiente a generare nelle persone-chiave il riconoscimento del mio essere sposato. Cioè, certe parole costituiscono un atto di dichiarazione ogni volta che il pronunciarle implica il mio essere effettivamente sposato. Di conseguenza, anche se io negassi la mia partecipazione al gioco linguistico del dichiarare, il fatto che due persone dichiarate sposate siano sposate rimane vero per definizione.

Ora dò un'interpretazione naturale all'atto della promessa: prometto qualcosa quando acquisisco il dovere di fare ciò che dico, e acquisisco il dovere di fare ciò che dico quando sono disposto a farmi punire da una società disposta a punirmi qualora non facessi quanto dico (se non si dà tale condizione, ho *detto* «prometto» ma non ho compiuto l'atto). Pigden può dunque rifiutare il gioco della promessa in due modi: o nega la sua esistenza e cioè afferma che io non devo fare quanto dico, o nega la sua partecipazione e cioè afferma che *lui* non farà quanto dice. Nel primo caso affermerebbe solo che non devo fare promesse (che non devo dire cose tali da doverle fare, cioè tali da suscitare in chi mi ascolta la disponibilità a punirmi), nel secondo caso invece affermerebbe solo che lui non farà promesse: in entrambi i casi quindi deve continuare a sostenere che chi fa promesse ha dei doveri. Potrebbe allora tentare di dimostrare l'impossibilità del gioco, ma anche così dimostrerebbe solo che è impossibile compiere l'atto della promessa, e dovrebbe continuare a riconoscere che *fosse possibile compiere l'atto* questo implicherebbe un dovere. Dunque è pressoché innegabile che l'atto della promessa implichi un dovere.

Pigden mi sembra fraintendere Searle due volte. Il primo fraintendimento sta nell'intendere la promessa solo come il *dire* (in un certo contesto sociale) di promettere, quando invece la promessa è il compiere un certo atto linguistico. Il secondo fraintendimento sta nell'aspettarsi che nella conclusione del ragionamento appaia il *verbo* «dovere», quando invece può apparire anche solo il *sostantivo* «dovere». Cioè, affinché si manifesti l'unità di fatti e valori, non è necessario che la conclusione sia «devo *x*», è sufficiente che sia la descrizione di un fatto costituente un dovere (come «tutti sono disposti a punirmi se non faccio quanto dico, e io non voglio essere punito»). In effetti credo che Pigden non sia disposto a definire *valori* dei doveri identici a certi fatti o stati di cose; eppure, tutte le teorie normative viste definiscono il valore come un fatto (piacere, rispetto delle leggi, virtù, eccetera).

È ovviamente possibile immaginare argomentazioni dello stesso tipo contro gli altri esempi, con simili risposte: un antirealista potrebbe dire «il capitano non ha il dovere di scendere per ultimo, è solo che viene punito se scende prima», il che presuppone che il dovere non sia il venire puniti se si fa diversamente, quando potrebbe proprio essere quello il significato della parola.

Le varie tesi 3: naturalismo e non-naturalismo

A questo punto sorge spontanea una domanda: se una affermazione come «il capitano deve scendere per ultimo» è vera, che tipo di fatto la rende tale? E cosa dire di affermazioni su funzioni e virtù? Nei quattro casi visti, ci sono fatti linguistici, sociali, legali, non soggettivi ma *intersoggettivi*: se l'umanità sparisse, sparirebbero anche i fatti che rendono vere tali affermazioni. Sono quattro casi la cui verità è evidente, ma, se si vuole formare un'etica, non possono essere gli unici veri: così fosse, allora un uomo che agisce al di fuori della sua professione potrebbe non avere più ruoli e funzioni, e sarebbe quindi sistematicamente immorale o almeno ingiudicabile in ogni sua azione; deve invece essere possibile rendere merito della verità di un'affermazione etica anche a prescindere da promesse, ruoli, funzioni, virtù epistemiche.

Infatti, i quattro casi hanno un elemento comune: rappresentano valori la cui validità è condizionale, nessuno punta ad un bene universale. Mi chiedo: è male scendere dalla nave prima di altri? Risposta: solo se sono il capitano (e lo sono in una società con una Marina strutturata in un certo modo). L'affermazione «è bene scendere per ultimi» è vera se la sto riferendo a certi individui, falsa se la sto riferendo ad altri, e che sia bene o male per me dipende da mie caratteristiche personali e dal mio contesto. Questo però non può essere valido per tutte le affermazioni etiche: io mi aspetto che «stuprare a morte bambini per divertimento è male» sia una verità universale, mi aspetto sia vero in tutte le circostanze a prescindere da qualsivoglia caratteristica personale, da credenze, emozioni, leggi e tutto ciò che è variabile.

Quali sono le caratteristiche di questi valori universali? In primo luogo, non possono costituire obblighi: se un obbligo fosse universale, allora, in ogni circostanza, dovrei ripetere infinitamente l'azione prescritta, se mangiare gelati fosse un obbligo universale dovrei passare ogni secondo della mia vita con un ghiacciolo in bocca. Possono però costituire dei divieti, possono presentare azioni che sono illegittime in tutti i contesti: non c'è un bene assoluto, ma un male assoluto sì.

In secondo luogo, proprio in quanto universali, il loro valore di verità deve rimanere identico anche al cambiare di ogni elemento intersoggettivo, di ogni struttura della società e della mente, al cambiare di ruoli e funzioni, alla formazione di nuove promesse (infatti qualcuno potrebbe promettere di fare un male assoluto, ma in tal caso il dovere che sorgerebbe dal promettere deve venire bloccato). Sono valori che, se sono veri, vogliono esserlo per ogni soggetto e ogni società, vogliono esserlo anche nel nulla cosmico dopo la morte termica dell'universo.

In terzo luogo, i valori universali devono riferirsi ad una esemplificazione (i filosofi dicono *istanziatura*) così piena e lampante della negatività che:

- ha tutte le caratteristiche di ciò che è riconoscibile come male
- e/o non ha alcuna delle caratteristiche di ciò che è riconoscibile come bene,
- quindi è inconfondibile con altre azioni che hanno alcune caratteristiche in comune con essa ma non sono male.

Le categorizzazioni sono spesso ambigue, sfumate. Per esempio, il concetto popolare di *uccello* potrebbe essere tale che, preso un qualsiasi ente, esso o è, o non è, un uccello; ciò nonostante, pare esservi un *continuum* tra uccelli e non-uccelli: il pinguino è un caso che può mettere in dubbio il parlante perché ha proprietà necessarie e non sufficienti all'essere uccello (ha il becco) ma anche proprietà *apparentemente* sufficienti all'essere non-uccello (nuota, non vola), potrebbe essere categorizzato nei *monotremi* assieme all'ornitorinco (si veda, e.g., Lakoff 1999). Questi però sono pensieri che non sorgono nei casi estremi: nessuno ha dubbi sul piccione che ha tutte e sole le

caratteristiche distintive dell'uccello, e nessuno definirebbe uccello l'ostrica che non ha nessuna caratteristica distintiva della categoria. Il male assoluto è estremo, è così diverso da tutto il bene che nessuno dei parlanti che comprendono il significato della parola «male» può essere in dubbio sulla categorizzazione.

Mi sembra che «stuprare a morte bambini per divertimento è male» possa rientrare in questa categoria.

Ciò che mi chiedo ora non è se questa affermazione sia vera, ma, se è vera, che tipo di fatto *potrebbe* renderla tale, e se questo fatto possa esistere considerando che non può essere una caratteristica della persona o della società, ma bensì deve essere qualcosa di comune a tutti gli stati possibili dell'universo, a tutti coloro che agiscono, a tutte le azioni scelte, qualcosa che permea la natura del cosmo. Nello specifico, devo chiedermi se può essere un fatto naturale, compatibile col naturalismo, o se invece deve essere un fatto non-naturale irriducibile a qualcosa di diverso dal male stesso. Non c'è dubbio che alcuni casi, per lo più imperativi ipotetici come il dovere di un capitano, siano resi veri da fatti naturali, ma devo chiedermi se ciò possa essere una regola universale valida per tutte le affermazioni etiche.

Non ci sono prove definitive 1: domande aperte e internalismo della motivazione

L'influenza di Moore nell'etica analitica si deve soprattutto alla sua argomentazione contro il naturalismo, l'argomentazione della *domanda aperta*. Moore (1903) tenta varie analisi del concetto di *bene*: bene come *felicità*, bene come *piacere*, bene come *i desideri che desidererei avere...* Secondo lui, però, quando l'analisi è corretta questo deve risultare subito evidente: l'analisi di *scapolo* è *uomo non-sposato* e ciò lo si può verificare nel fatto che la domanda «lui è scapolo, ma è anche un uomo non-sposato?» suona assurda, tautologica, è una *domanda chiusa* perché la sua risposta sta nel significato dei termini che la compongono. Se faccio la prova e mi chiedo «quest'azione è piacevole, ma è anche bene?» non ho la stessa sensazione, è una *domanda aperta* la cui risposta va scoperta indagando il mondo. Posso ripetere la prova con varie analisi, posso provare con tutte le proprietà naturali, e così vedo che nessuna costituisce un'analisi corretta: quindi il bene è irriducibile a fatti naturali e, in generale, a qualsiasi fatto non-etico.

Questa argomentazione, all'inizio del Novecento, era molto forte; oggi, con le nuove conoscenze prodotte dalla logica e dalla filosofia del linguaggio, non convince più. Io posso chiedermi «questa sostanza è acqua, ma è anche H₂O?» e questa è una domanda aperta, eppure è vero che l'acqua è H₂O; similmente, chiedermi se il bene sia il piacere costituisce una domanda aperta, ma questo non implica che il bene non sia davvero il piacere, implica che l'equivalenza tra bene e piacere non è analitica e vera a priori, ma potrebbe essere scoperta a posteriori, potrebbe essere un caso di *identità sintetica* tra due proprietà (cfr. Brink 2001).

Coloro che oggi difendono l'argomentazione di Moore lo fanno aggiungendo un elemento che permette di spiegare *perché* le domande risultano aperte, cioè perché è impossibile ridurre le proprietà etiche a proprietà naturali: la *motivazione morale*.

È credenza comune, sin da Platone (così Mackie 2001, cfr. l'Apologia di Socrate), che i fatti etici siano *intrinsecamente motivanti*: il mio sapere che uccidere è male è sufficiente a motivarmi a non uccidere, il mio sapere che fare la carità è bene è sufficiente a motivarmi a farla (anche se magari con una forza insufficiente a farmela attuare). Ciò che distingue questo tipo di motivazione è la sua

indipendenza dal *desiderio*: posso avere il desiderio di uccidere, ma sono comunque motivato a non farlo; posso non avere il desiderio di fare la carità, ma sono comunque motivato a farla. Nessun fatto naturale, però, pare avere questa capacità: il fatto che il mondo sia strutturato in un certo modo è cosa neutrale, non è sufficiente a darmi alcuna motivazione, qualsiasi sia la natura del mondo io posso sempre rispondere «e allora?»; quindi i fatti etici intrinsecamente motivanti devono essere non-naturali.

Questa interpretazione della motivazione costituisce l'*internalismo* in etica. Hume (1739), come corollario della sua distinzione tra fatti e valori, sposa invece l'*esternalismo*: secondo lui il sapere che una certa azione sia bene è insufficiente a motivare, è necessario che vi sia anche il desiderio di compiere quell'azione. L'esternalismo è coerente con l'esistenza di azioni tali da essere impossibile il non desiderarle, tali da essere sempre accompagnate da desiderio; la loro capacità di motivare, però, dipende da questo, se non ci fosse il desiderio non sarebbero neanche motivanti (si vedano Williams 1981, McDowell 1998, Copp 2001, Savarivsdottir 1999).

Un internalismo globale è senza dubbio da rifiutare: infatti, se l'internalista vuole difendere il non-naturalismo, e accetta che il dovere del capitano di restare sulla nave emerga da un fatto naturale, allora non può affermare che il capitano sia intrinsecamente motivato a restare sulla nave, perché in tal caso un fatto naturale sarebbe sufficiente a motivare. Al massimo, allora, può sostenere che i valori universali (e non tutti) siano intrinsecamente motivanti. Epperò, non pare esserci alcuna contraddizione in un pedofilo sadico che affermi «so che stuprare a morte bambini per divertimento è un male, ma non mi interessa minimamente, proprio zero»: se è così, allora, qualsiasi sia il fatto su cui si basa quel valore, non è intrinsecamente motivante.

Anscombe (2016) e Searle (2003, 2009), riprendendo una tradizione medievale, difendono un tipo di esternalismo tramite la *direzione di adattamento*: la credenza etica va dal mondo alla mente, il desiderio etico va dalla mente al mondo; la credenza sta nell'adattare la mente al mondo per descriverlo ovverosia ha il suo oggetto nel mondo com'è, il desiderio sta nell'adattare il mondo alla mente per realizzare ciò che si immagina ovverosia ha il suo oggetto nel mondo come deve essere. Ne consegue che una credenza etica, proprio perché credenza, non può essere intrinsecamente motivante. Posso obiettare che credenze e desideri non sono mai completamente divisi: io so di avere *meta-desideri* e *meta-credenze*, ho credenze sulle credenze e desidero avere certi desideri, ho credenze sui desideri e desidero avere certe credenze. Questo, però, non sfata l'esternalismo, al contrario lo conferma: proprio perché credenze e desideri sono inseparabili la credenza senza desiderio deve essere insufficiente a motivare.

È questo un problema per il non-naturalista? Non pare: non c'è contraddizione nel sostenere che il bene è una proprietà non-naturale e non-motivante; di contro, il naturalista pure potrebbe sostenere che ci sono proprietà naturali e motivanti (in quanto irresistibilmente desiderabili). Così, sia internalismo che esternalismo possono essere coerenti sia con il naturalismo che con il non-naturalismo, quindi l'internalismo non può essere un'argomentazione contro il naturalismo, né può difendere l'argomentazione della domanda aperta.

Non ci sono prove definitive 2: supervenienza e conoscenza

Ora mi domando: è possibile immaginare un universo totalmente identico a questo nei fatti fisici, ma diverso nei soli fatti etici? È possibile che il giudizio etico su un'azione in un certo contesto cambi senza che cambi qualcosa nel contesto?

Tra gli oggetti possono esservi varie relazioni: «essere uguale a», «stare alla destra del», eccetera. Una di queste relazioni è quella di *supervenienza*. Dire che un certo ente *x* *superviene* su un altro *y* vuol dire che «affinché cambi *x* è necessario che cambi *y*»: per esempio, se la mente superviene sul cervello, allora la mente può cambiare solo se cambia qualcosa nel cervello (mentre il cervello può cambiare senza che cambi la mente), quindi se ho un'esperienza mentale e cioè cambia qualcosa nella mente so che deve essere avvenuto un qualche evento neurale che ha cambiato qualcosa nel cervello (Kim 1993).

La nozione di supervenienza, che ora è applicata in tutti i campi della filosofia, è nata proprio in etica (forse con Hare 1952): tra gli eticisti, infatti, è molto comune l'intuizione che i fatti etici debbano supervenire sui naturali. È un'idea facile da supportare. In primo luogo, l'etica guida l'azione *nel* contesto, i principi etici mi dicono che se si dà un certo caso allora devo agire in un certo modo; ma se invece i valori etici cambiano in modo indipendente dal contesto, allora l'etica non può più essere una guida per l'azione (Blackburn 1993). In secondo luogo, l'idea che, in un identico contesto, possa essere bene sia un'azione, sia il suo contrario, va contro il *principio di non-contraddizione*: se «uccidere è male» è vero, non può esserlo anche «uccidere non è male», in un identico contesto si deve dare o uno o l'altro. In terzo luogo, se il bene superviene su fatti naturali come comportamenti o leggi, allora io posso spiegare cos'è il bene anche a chi non lo conosce appellandomi ai fatti del contesto; ma se non è così, e il bene può cambiare in modo indipendente, allora comunicare cos'è il bene diventa quasi impossibile, e quindi dovrei mettere in dubbio la mia stessa comprensione del concetto. Perciò: un universo uguale a questo in tutti i fatti naturali deve essere identico a questo anche nei fatti etici, se un'azione è bene in uno lo è anche nell'altro, e se invece c'è una differenza etica allora deve esservi anche una differenza naturale, ci sono differenze naturali senza differenze etiche ma non viceversa.

Perché la supervenienza è rilevante per il naturalismo? Perché, se i fatti etici *sono* fatti naturali, allora tra i due *deve* esservi supervenienza, è impossibile che uno cambi senza l'altro in quanto sono la stessa cosa; ma se invece i fatti etici sono *sui generis*, se sono indipendenti, allora il motivo per il quale uno cambia solo al cambiare dell'altro rimane un mistero inspiegato (Dreier 1993).

Questa può essere un'argomentazione pragmatica efficace, ma l'idea che fatti naturali ed etici supervengano senza essere identici resta logicamente coerente. Per esempio, se Dio è un ente non-naturale, e si sostiene una morale religiosa, allora in due universi fisicamente identici Dio potrebbe comandare cose diverse, di modo che, a parità di fatti naturali, vi sarebbero diversi fatti etici. Se non si dà questo caso e fatti naturali ed etici supervengono è perché Dio, essendo un ente razionale, trova inaccettabile che un principio che deve guidare l'azione in modo logico vari in modo indipendente dal contesto: il bene deve formare un tutt'uno col contesto, cioè deve essere il contesto a determinare quali azioni sono possibili e le azioni devono modificare tale contesto, ma il bene rimane distinto da tutto il resto.

Posso allora puntare il dito al fatto che, supervenendo, un'azione è bene ogni volta che di danno certe proprietà naturali, e non è mai bene quando queste non ci sono: ciò può dimostrare che il bene è identico a quelle proprietà naturali. Epperò non è necessario sia così: ciò che posso fare è individuare una equivalenza logica tra proprietà etiche e naturali, per esempio posso notare che tutto ciò che è bene è piacevole e tutto ciò che è piacevole è bene; questo, però, è insufficiente al dimostrare l'identità tra le proprietà. Qualche devoto amante della natura potrebbe dire che la grandiosità delle montagne implica l'esistenza di Dio, e l'esistenza di Dio implica che le montagne debbano essere grandiose: ciò crea una equivalenza logica tra la grandiosità delle montagne e l'esistenza di Dio, il conoscere uno permette di conoscere l'altro, ma sarebbe assurdo ritenerle proprietà identiche.

I valori universali, infine, rendono merito della supervenienza senza implicare identità: un'affermazione come «mentire è male» è tale che il suo valore di verità cambia col contesto, cambia al variare dei fatti naturali (mentire per proteggere un perseguitato non è male); ma l'affermazione «stuprare a morte bambini per divertimento è male» è insensibile alle variazioni del contesto, in qualsiasi modo cambino i fatti fisici il valore etico non cambia. Ciò è coerente con la relazione di supervenienza, ma spinge a negare l'identità: infatti, se tutti i fatti naturali possono cambiare senza che il fatto etico cambi, questo deve essere diverso da tutti i fatti naturali. D'altronde, non c'è motivo per il quale un valore etico universale dovrebbe variare insieme a fatti intrinsecamente non-etici, devono essere irrelati.

Quindi: la supervenienza è insufficiente al dimostrare il naturalismo, e il non-naturalista può spiegarla senza appellarsi all'identità tra fatti etici e naturali, così entrambe le teorie rimangono valide. Il naturalista ha bisogno di una strategia diversa.

Posso concentrare gli attacchi contro il non-naturalismo sull'aspetto epistemologico. Un'altra idea popolare tra epistemologi ed eticisti è questa: è ingiustificato credere all'esistenza di un fatto inosservabile se questo non spiega qualcosa di osservabile; cioè, il fatto inosservabile può essere creduto giustificatamente solo se abdotto da fatti osservabili. Esempio: osservo fatti come fulmini e reazioni chimiche, nessun altro fatto osservato spiega il perché e il percome di questi, gli elettroni sono inosservabili ma sono definiti da proprietà tali da implicare quei fatti, quindi dall'esistenza dei fulmini e vari altri fatti fisici, chimici, biologici posso aburre l'esistenza dell'elettrone. Esempio contrario: nella foresta osservo enormi impronte e rami alti rotti, il Big Foot (che nessuno ha mai osservato) è una possibile spiegazione, ma ci sono molti altri fatti osservati (animali, o umani ingannatori..) che possono spiegare ciò che vedo, quindi sarebbe illegittimo aburre il Big Foot come causa. Ora, vi fossero fatti etici, sarebbero inosservabili, sarebbero *queer*, strani, diversi da tutto ciò che la scienza può descrivere, e non sono necessari a spiegare alcunché (Mackie 2001): è possibile affermare che le rivolte contro un regime sono causate dalla *crudeltà* del dittatore (e.g. Sturgeon 1988), ma è un'ipotesi illegittima perché posso spiegare la rivolta tramite processi psicologici e sociali naturali. Quindi, per quanto ne posso sapere, non ci sono fatti etici non-naturali, i valori universali o sono naturali o non esistono.

Un'altra critica epistemologica riguarda il metodo di conoscenza dell'etica. Se i fatti etici sono fatti naturali allora non c'è alcun problema: ricordo che anche Wittgenstein, nel suo periodo più pragmatista, ammette che l'etica si scopre come qualsiasi altra cosa, con lo stesso metodo. Ma se non è così? Se esprimo una proposizione conoscibile a posteriori come «in Calabria il cielo è verde» so subito come controllare il suo valore di verità, devo andare in Calabria e osservare il cielo. Se esprimo una proposizione conoscibile a priori come «il rosso è un numero» capisco subito il suo valore di verità in virtù del significato delle parole. Ma un'affermazione come «stuprare bambini è male» non è conoscibile né a priori, né è a posteriori in quanto, cercando per ipotesi di riferirsi ad un fatto non-naturale, non può essere controllata osservando il mondo, né esiste un *metascopio* che permetta di osservare i fatti etici non-naturali.

Come mai è così? I fatti naturali sono conoscibili a posteriori perché *causalmente interagenti*, i fatti non-naturali invece non hanno potere causale: un sasso, interagendo con la luce, può *causare* il mio vederlo e pensarlo, e ciò giustifica la mia credenza sulla sua esistenza; ma un fatto etico non-naturale non può causare pensieri o percezioni, quindi, anche vi fossero fatti non-naturali, non potrei mai averne conoscenza, resterebbero rinchiusi nel loro mondo inaccessibile (Stratton-Lake 2020). Ne consegue che, quando io esprimo un giudizio morale su un'azione, solo un fatto naturale può aver causato quel giudizio; quindi ci sono due possibilità: o il valore etico sta in un fatto

naturale, e ho acquisito conoscenza etica; o sta in un fatto non-naturale inconoscibile, e quindi qualsiasi giudizio etico è ingiustificato, risultando in scetticismo.

Detto tra parentesi: l'epistemologia intuizionista

Ho evidenziato due problemi epistemologici: la necessità di abdure gli inosservabili dagli osservabili, e l'impossibilità di conoscere fatti non-naturali senza potere causale. La risposta dei non-naturalisti è diretta: non è vero, né l'abduzione né il potere causale sono necessari alla conoscenza.

C'è un intuitivo problema epistemologico, il *problema delle altre menti*. Io so di non essere l'unico umano cosciente al mondo, è in assoluto la mia intuizione più forte, io so che tutti gli esseri umani hanno una mente, so che ognuno ha esperienze soggettive, mi è indubitabile. Epperò tali esperienze sono per definizione inosservabili, solo chi le vive ne ha coscienza; per di più, non sono necessarie a spiegare alcunché, tutto ciò che è osservabile in una persona può essere spiegato tramite i processi chimici del suo corpo, se gli esseri umani non avessero esperienze coscienti potrebbero cionondimeno fare tutto ciò che fanno come dei robot (Plantinga 1967). Com'è possibile allora avere conoscenza dell'esistenza delle altre menti? È un gran dilemma, ma l'idea che l'esistenza di altre menti sia inconoscibile, che sia solo un'ipotesi valida tanto quanto l'idea che io sia l'unico essere cosciente, è inaccettabile. Conseguenza: non è necessaria l'abduzione per conoscere un fatto inosservabile, infatti la mente è sia inosservabile che non-abducibile legittimamente da fatti osservabili, ma resta conoscibile. Com'è possibile conoscerla? Potrebbe essere dedotta da altre credenze, o essere inserita in un sistema per aumentarne la coerenza, comunque tali metodi non richiedono alcun potere causale.

Per gli *intuizionisti* (e.g. Moore 1903) il modo in cui si conosce un fatto etico è simile a questo, è simile al modo in cui si conosce un colore. Il rosso di per sé non spiega nulla, e non posso neanche spiegare ad altri cosa esso sia, potrebbe esistere un universo con solo lunghezze d'onda, senza colori così come sono percepiti; eppure li percepisco, e ciò è sufficiente a intuire la loro esistenza. Similmente, quando penso ad un omicidio, percepisco subito che è un male, è questo è tutto ciò di cui ho bisogno per fare conoscenza del male.

Questa *esperienza etica* del bene e del male ha due aspetti, uno emotivo e uno razionale. Dal punto di vista emotivo, noto che l'esperienza è intrinsecamente motivante proprio come vogliono gli internalisti: se metto la mano sul fuoco, il sentire dolore è sufficiente a motivarmi a toglierla; similmente per il male, il fatto che io conosca l'etica tramite percezione spiega perché sono intrinsecamente motivato a seguire o evitare certi comportamenti. Dal punto di vista razionale, invece, dico che il fatto etico, pur non comunicabile ad altri, è magnificamente presente innanzi alla mia ragionevolezza: per me, negare che l'omicidio sia male è come negare che la mela che ho qui davanti sia rossa, inconcepibile (cfr. Huemer 2005).

Eppure il giudizio etico non sembra paragonabile alla percezione di un oggetto. Infatti, quanti sono in disaccordo sul fatto che il cielo sia blu? Quasi nessuno, e chi lo è ha delle patologie, è daltonico. Quanti sono in disaccordo sul fatto che il David di Michelangelo sia bello, che produca sensazioni associate alla bellezza? Pochi, ma comunque un po' di più di quelli che pensano che il cielo non sia blu. E quanti sono in disaccordo sul fatto che l'aborto sia un male? Tantissimi, è un argomento controverso che divide la popolazione in schieramenti. Se conoscere l'etica è come riconoscere un colore, perché c'è un così diffuso accordo sui colori, e un così diffuso disaccordo

sulle questioni etiche? Se ognuno può intuire direttamente i fatti etici, perché non c'è accordo universale? (Stevenson 1962, Gibbard 1990)

Una risposta è che in tutto c'è disaccordo, ma sulle questioni etiche in modo particolare perché sono complesse, hanno più elementi coi quali si può essere in disaccordo: chi crede che il cielo non sia blu, lo crede solo perché non lo vede così; ma chi crede che l'aborto non sia un male potrebbe crederlo perché pensa che l'omicidio sia bene, o perché pensa che un feto non sia una persona, o perché pensa che in certi contesti sia abortire sia il non farlo possano manifestare diverse virtù senza contraddizione, eccetera.

Secondo alcuni ciò è insufficiente a rendere merito dell'ampiezza del disaccordo etico, e bisogna riconoscere che la sua origine sta in un'assenza di conoscenza (Campbell 2019); secondo altri, come Donald Davidson, in realtà il disaccordo etico non è così ampio come si può pensare. Davidson (1994) segue l'idea che non si possa cambiare il valore di un termine spesso senza cambiarne il significato: se penso che rubare sia bene non posso usare la parola «onesto» con lo stesso significato che gli altri parlanti le danno, quindi maggiore è il disaccordo etico, maggiore è la confusione sul significato dei termini spessi; eppure è evidente che il significato dei termini spessi è condiviso, e che i concetti spessi possono essere tradotti e comunicati anche a società diverse dalla propria, quindi non può esservi un gran disaccordo etico.

L'argomentazione di Davidson, però, non riesce a spiegare il fatto che, nonostante il significato dei miei termini spessi sia stabile, di fatto le mie convinzioni etiche cambiano molto di frequente. Se la mia stessa idea di *bene* evolve nel tempo (come invece non fa la mia idea di *rosso*), allora questa non può essere un'intuizione o una percezione, deve essere invece un ragionamento complesso, diretto a fatti che sono difficili da scoprire.

Non ci sono prove definitive 3: la ramseizzazione

Il genio di Frank Ramsey ha lasciato un'ampia eredità alla filosofia: dopo la teoria deflazionaria della verità, ecco ora la *ramseizzazione*, il *metodo di Ramsey-Lewis* per l'analisi dei termini (Ramsey 1964, Lewis 1970).

Esempio. Ho sottomano una teoria fisica dell'elettrone, una teoria che contiene varie affermazioni su di esso e sulle sue proprietà. Tale teoria, però, non definisce esplicitamente il termine «elettrone», e trattandosi di un ente inosservabile nessuno ha nozione intuitiva della sua natura: come posso sapere, quindi, a cosa dovrebbe riferirsi la tal parola? L'idea del metodo di Ramsey-Lewis è che le affermazioni della teoria dell'elettrone contengano implicitamente la sua definizione, la quale determina l'estensione, cioè l'insieme di enti che rendono vera l'affermazione «questo è un elettrone». Nello specifico, l'elettrone è quell'ente che può rendere vere tutte le affermazioni della teoria.

Immagino allora di avere questa teoria:

- gli elettroni hanno carica negativa; ciò che ha carica negativa può mandare scosse; Feynman ha un elettrone in tasca.

Una volta ramseizzata diventa così:

- esistono le proprietà $P1$ e $P2$ tali che: ciò che ha la proprietà $P1$ ha anche la proprietà $P2$; ciò che la proprietà $P2$ può mandare scosse; Feynman ha in tasca un ente con la proprietà $P1$.

$P1$ indica l'*elettronità*, l'essere elettrone (e $P2$ l'avere carica negativa). La definizione implicita di elettrone allora è «ciò che ha la proprietà $P1$ » dove $P1$ è tale da soddisfare la frase ramseizzata (Toraldo di Francia 1976).

Come Ramsey, anche Frank Jackson è esperto sia di filosofia che di matematica; è lui che ha avuto l'intuizione di applicare il metodo della ramseizzazione ai termini etici, in particolare a «bene». Per poterlo fare c'è bisogno di una teoria etica di partenza: quale scegliere? Jackson (1998) mette insieme delle *platitudini*, affermazioni trite e ritrite, proverbiali, ma considerate incontrovertibili, condivise da tutti. Per esempio, posso formulare questa teoria del bene:

- ciò che è bene va incentivato; le persone virtuose hanno una tendenza al bene.

Quindi la ramseizzo:

- esistono le proprietà $P1$ e $P2$ tali che: ciò che ha la proprietà $P1$ va incentivato; chi ha la proprietà $P2$ ha la tendenza a fare ciò che ha la proprietà $P1$.

Il bene deve essere un'unica proprietà, non posso usare questa frase come sua definizione in quanto la proprietà $P1$ potrebbe in realtà indicare un insieme di altre proprietà diverse; quindi devo specificare l'unità e unicità di $P1$ così:

- il bene è la proprietà $P1$, la quale è tale che: esiste la proprietà $P2$ tale che ciò che ha la proprietà $P1$ va incentivato e chi ha la proprietà $P2$ ha la tendenza a fare ciò che ha la proprietà $P1$; per tutte le altre proprietà $P3$ e $P4$, ciò che ha la proprietà $P3$ va incentivato e chi ha la proprietà $P4$ ha la tendenza a fare ciò che ha la proprietà $P3$ sse $P3 = P1$ e $P4 = P2$.

Sia chiaro che il fatto che $P1$ sia un'unica proprietà non vuol dire che sia una *proprietà semplice*, può invece essere *complessa*. Un esempio è il concetto di *salute*: che una persona sia salubre non è qualcosa di riconducibile ad un unico aspetto, la salute non è solo fisica o mentale, e non è neanche qualcosa di osservabile in tutti i casi, posso vedere una persona d'aspetto salutare ma non posso escludere che in realtà abbia qualche malattia; la salute è un complesso insieme di caratteristiche, solo a volte inferibile da certe proprietà (l'aspetto, la vitalità) che sono causate dalla salute pur essendo diverse da essa. Il bene, come anche la virtù, è una proprietà di questo tipo: complessa, non sempre osservabile, inferibile da ciò che causa (cfr. Boyd 1988).

Ora, nell'esempio della Terra* gemella la parola «acqua» indica ciò che, in un certo mondo, svolge la *funzione* dell'acqua: sul mio pianeta la sostanza H_2O riempie i fiumi e i laghi, può essere usata per fare il gelato, ogni tanto cade dal cielo, eccetera; se, su un altro pianeta, una diversa sostanza XYZ costituisce i fiumi, i gelati, le nuvole, allora svolge la stessa funzione di H_2O e quindi è acqua. Altro esempio: cos'è un cuore? È ciò che pompa il sangue: il fatto che abbia quattro cavità, o due, o sia un vaso pulsante, o abbia la forma di un orsetto, è irrilevante, se svolge la funzione del pompare sangue è un cuore (cfr. Putnam 1987). Jackson fa lo stesso ragionamento: la proprietà $P1$ ha un certo ruolo in questa rete di affermazioni, tale è la sua funzione, e qualsiasi cosa la svolga quella è il bene. E secondo Jackson questa proprietà deve essere naturale, sempre a causa della supervenienza: poiché le proprietà etiche non possono cambiare senza che cambi qualche

proprietà naturale, la proprietà *P1*, se etica, sarà sempre logicamente equivalente a qualche proprietà naturale.

Il problema dell'argomentazione di Jackson è che esclude l'esistenza di proprietà distinte ma *coestensive*. Cioè, per lui, se tutto ciò che è bene è (per esempio) anche tutto ciò che è piacevole, il bene è identico al piacere, indistinto dal piacere. Eppure, la proprietà dell'avere tre angoli è coestensiva alla proprietà dell'avere tre lati, e restano proprietà distinte.

Se salta questo punto, la ramseizzazione di per sé non permette di scegliere tra naturalismo e non-naturalismo. È come se mi chiedessi «un cuore è organico?»: se il cuore è definito da una funzione, no, perché ci sono anche macchine che svolgono la stessa funzione di un cuore; similmente, se dò una definizione funzionalista di «bene», alla domanda «il bene è naturale?» potrò rispondere di no, perché, anche se con quelle platitudini io posso volermi riferire a proprietà naturali, comunque potrei altrettanto legittimamente riferirmi a proprietà non-naturali che svolgono la stessa funzione in quella rete di concetti.

Non ci sono prove definitive 4: ancora Terre gemelle*

Horgan e Timmons, eticisti americani, prendo ispirazione da un'argomentazione di Putnam per difendere il non-naturalismo metaetico.

Anzitutto, devo presupporre la *teoria causale del riferimento* per le *sostanze naturali*: se davanti ad un bicchiere di H₂O dico «acqua», e lo stesso dice l'uomo* della Terra* davanti ad un bicchiere di XYZ, i nostri riferimenti sono rispettivamente H₂O e XYZ perché quelle sono le sostanze che hanno *causato* il nostro comportamento, sono le sostanze che *regolano l'uso* della parola nei due casi (cfr. Kripke 1999, Putnam 1987).

In tale teoria, in primo luogo, una volta battezzata una sostanza come «acqua» io posso indagarla e, qualsiasi proprietà scoprirò, potrò accettarla come proprietà dell'acqua, cioè nessuna proprietà potrà farmi dire «mi sono sbagliato, non era acqua». In secondo luogo, se io incontro l'uomo* della Terra* e noto un contrasto nel nostro uso di «acqua», cioè noto che usiamo gli stessi fonemi per riferirci a sostanze diverse, concludo che stiamo usando parole diverse (pur con lo stesso suono), e non che uno dei due non ha compreso la parola.

Ora immagino un caso analogo, ma con la parola «bene». Arrivo sulla Terra*, e scopro che gli abitanti utilizzano la parola «bene» in modo simile al mio, la usano per presentare azioni da incentivare e giudicate degne delle persone virtuose. Scopro anche che la usano per descrivere le azioni di chi porta da mangiare ai poveri e compie altri atti di carità. Indagando quell'azione, però, scopro che su quel pianeta si dà ai poveri solo cibo scaduto per far loro dispetto.

In primo luogo, non posso più accettare che tale azione sia un bene, l'avevo definita tale ma ho scoperto una caratteristica che mi ha fatto cambiare idea, e ciò implica che l'uso del termine «bene» non può essere causato o regolato dall'azione che descrive. In secondo luogo, se gli abitanti della Terra* insistono nel dire che quell'azione è «bene», devo concludere che sono moralmente devianti, e non che usano un concetto etico diverso dal mio (altrimenti il disaccordo etico sarebbe impossibile anche tra diverse società terrestri).

Il comportamento di «bene» è nettamente distinto da quello di un termine naturale come «acqua». Perché non risponde alla medesima semantica? La risposta diretta è: perché non è un termine naturale, indica qualcosa di non-naturale. Infatti, se quella esposta è realmente la semantica di «bene», le argomentazioni tradizionali a favore del naturalismo crollano (Horgan & Timmons 1991).

Mentre posso scoprire qualcosa in più sull'acqua indagando la sostanza che chiamo con quel nome, se l'uso del termine «bene» non è regolato dalle azioni che caratterizza, non posso scoprire qualcosa in più sul bene analizzando le azioni che definisco tali. Questo fa crollare le argomentazioni epistemologiche: se per il non-naturalista è difficile spiegare come sia possibile conoscere fatti non-naturali, il naturalista non ha gioco più facile, anche lui deve spiegare come sia possibile conoscere una proprietà delle azioni senza poter indagare le azioni medesime.

Anche l'argomentazione tramite supervenienza crolla: se l'uso del termine «bene» fosse regolato dalle caratteristiche naturali degli eventi che descrive, allora il perché l'uso del termine dipenda da fatti naturali, perché non possa cambiare la descrizione etica senza che cambi la descrizione fisica, sarebbe ovvio, ma poiché l'uso del termine è indipendente dalle caratteristiche naturali dell'azione, sì, può essere vero che il non-naturalista manchi di una spiegazione per la supervenienza, ma lo stesso vale per il naturalista.

Infine, l'argomentazione della domanda aperta di Moore ne risulta rafforzata: Moore si limita a dimostrare che non è stata ancora scoperta una equivalenza analitica tra il bene e una qualche proprietà naturale, l'argomentazione di Horgan e Timmons in più mostra che anche una equivalenza sintetica a posteriori è impossibile, se i termini etici fossero equivalenti ad altri termini naturali mostrerebbero lo stesso comportamento semantico ma così non è (cfr. Huemer 2005).

Ciò nonostante, la conclusione è frettolosa. Horgan e Timmons non attaccano il naturalismo, attaccano il *realismo di Cornell* (un naturalismo basato sulla teoria causale, e.g. Boyd 1988) nella speranza di riportare in auge teorie non-cognitivistiche. Ma in effetti, in epistemologia, ho già detto che la teoria causale del riferimento è insufficiente, e anche la si accettasse non è necessario che tutti i termini indicanti sostanze naturali rispondano alla stessa semantica. Il fatto che «bene» non risponda alla semantica di «acqua», lungi dal dimostrare il non-naturalismo o il non-cognitivismo, potrebbe dimostrare invece che ci sono sostanze naturali che seguono semantiche diverse.

Volendo, posso ipotizzare giustificazioni per queste differenze semantiche. Per esempio, il motivo per il quale non posso scoprire qualcosa sul bene indagando le azioni che hanno la proprietà di essere bene potrebbe stare nel fatto che il termine «bene» ha un contrario, il «male», che è diverso dal «non-bene»; «acqua» invece non ha un contrario di questo tipo. Di conseguenza, quando dico che un'azione è «bene» dico anche che è «non-male», dico non solo che ha certe caratteristiche ma anche che *non* ne ha alcune, e non posso accettare come bene ciò che scopro avere caratteristiche in comune col male; al contrario, quando dico che una sostanza è «acqua» intendo solo dire che ha certe caratteristiche osservate, non è necessario che sia priva di altre proprietà, quindi non c'è limite a ciò che posso scoprire. Il definire un'azione «bene» non può essere un *battesimo* perché devo rimanere aperto alla possibilità di rivedere il mio giudizio nel corso dell'indagine.

Il motivo per il quale non sono disposto a risolvere un conflitto nell'uso di «bene» dicendo che io e l'uomo* della Terra* stiamo usando parole diverse può stare invece nel fatto che «bene» è un termine *valutativo* e «acqua» *descrittivo*. Io posso comunicare ad altri il significato di «acqua» indicando la sostanza della quale voglio parlare: questo crea l'estensione del termine, indagando poi gli enti dell'estensione costruisco l'intensione cioè la definizione. «Bene» è diverso: indica

un'astrazione condivisa da più azioni, azioni che hanno sia la proprietà del bene sia altre, non indica un ente nella sua interezza come «acqua». Conseguenza: non posso definire «bene» indicando qualcosa, devo partire da una *definizione operativa* (che è anche l'*obiettivo* dell'uso della parola) e solo dopo vado a individuare per tentativi le azioni che rientrano nell'estensione. Quindi, il fatto che un abitante della Terra* indicando la sostanza XYZ produca un'estensione diversa di «acqua» è sufficiente a cambiare anche l'intensione e così creare una parola diversa; ma in «bene» l'intensione precede l'estensione, l'indicare un'azione malvagia come «bene» non crea una nuova parola perché non genera una nuova intensione. Fintanto che usiamo «bene» con uno stesso scopo, cioè fintanto che la usiamo per indicare azioni da incentivare, stiamo usando tutti la stessa parola.

Infine, si può rifiutare la base della dell'argomentazione: secondo Andrea Viggiano (2008) il funzionamento dei termini etici in Horgan e Timmons è perfettamente coerente col funzionamento dei termini naturali in Putnam, quindi non è vero che l'esperimento mentale della Terra* gemella porta a conclusioni diverse per i due termini. Le diverse conclusioni alle quali giungono Horgan e Timmons possono essere in parte attribuite alla diversità del loro stile filosofico rispetto a Putnam: mentre Putnam espone il suo esempio limitandosi ai dati comportamentali osservabili, loro sono più liberali e presuppongono la conoscenza di elementi ulteriori.

I fatti vengono individuati a partire dai valori 1: una teoria putnamiana

Io mi aspetto che «stuprare a morte bambini per divertimento» sia un'azione inaccettabile per chiunque in qualsiasi circostanza. Se è così, posso aspettarmi anche che l'inaccettabilità dell'azione sia intrinseca al significato dei termini, cioè che l'universalità del valore sia spiegata col fatto che l'affermazione «stuprare a morte bambini per divertimento è male» sia analitica, il soggetto «stuprare a morte bambini» deve contenere già il «male» predicato. Può essere così? Per capirlo, dovrei analizzare i termini. Mi chiedo allora: cosa vuol dire «bambino»? Il significato della parola, in questo caso, non è ciò che io intendo (lo so già, cosa intendo), è un qualcosa di determinato socialmente, quindi: cosa intende la società con «bambino»? Le argomentazioni di Quine e Kripke mostrano che non c'è un'unica risposta, ci sono infinite interpretazioni per me legittime. Ciò che concludono è che l'assegnazione del significato alla parola non è una questione di fatti, non c'è un fatto in merito, è una questione pragmatica. Putnam, io credo, fa la specificazione ulteriore: se non è una questione di fatti, è almeno una questione di valori.

È un'affermazione radicale. I ragionamenti di Quine e Kripke non riguardano solo le parole, ma l'*interpretazione del mondo* in generale. Posso parlare del modo in cui collego una parola ad un fatto per determinarne il significato, ma questo è identico al parlare, per esempio, del modo in cui interpreto una percezione: come «gavagai» può essere interpretato sia come *coniglio* che come *parti non-distaccate di coniglio*, così la mia esperienza di una mela può essere interpretata come una mela reale, come un'allucinazione, come l'esperienza di un ente ideale, come l'esperienza di un gatto, eccetera. Cioè, la mia esperienza della mela in assenza di una *teoria della mela* non mi dice niente sul mondo, anche solo il mio essere in grado di distinguere una mela dal tavolo che la regge (invece di affermare che costituiscono un unico oggetto) è questione d'interpretazione. E se l'interpretazione è una questione di valore e non di fatti, tutto questo lo è, non solo il significato delle parole. In che senso può essere così?

Quando dico che «gavagai» vuol dire *coniglio* anziché *parti non-distaccate*, quando dico che il significato del termine va determinato pragmaticamente e non lanciando una moneta, quando

individuo una mela distinta dal tavolo sul quale poggia, quando interpreto la mia percezione come una mela reale e non come un'entità ideale, insomma ogni volta che applico i miei schemi mentali sul mondo, sto *preferendo* una concettualizzazione rispetto ad altre possibili; e in quanto preferenza, è una manifestazione dei valori.

Per esempio, in quanto scritto finora, io ho sempre ragionato in accordo con una certa logica, io accetto che se una cosa è rossa non può essere anche non-rossa: ma perché dovrebbe essere così? Ho sempre sostenuto l'esistenza di fatti, l'esistenza di un mondo oggettivo: perché non lo metto in dubbio? Perché non posso pensare che la realtà funzioni con categorie molto diverse da quelle che posso concepire, e che la categoria dei fatti sia irreali? Risponderei: dato che nessun ragionamento può avere una profondità infinita, devo pur partire da una base, e questa base, dovendo precedere ogni altra possibile discussione sui fatti, non potendo cioè dipendere a sua volta da fatti, può essere determinata solo da valori, i miei valori implicano l'esistenza di certi fatti descritti da una certa logica.

In altre parole: i fatti, se esistono, sono indipendenti da tutto ciò che io posso pensare, ma il mio essere in grado di *individuare* fatti, e quindi credere nella loro esistenza, è questione d'interpretazione; gli atomi esistono a prescindere dall'etica, ma io non posso dire «*so* che esistono atomi» in modo indipendente dal valore: se non vi fossero dei valori, io non potrei individuare alcun fatto (perché non avrei modo di favorire un criterio d'indagine sugli altri), non potrei dire cosa significa la parola «fatto» e quindi non potrei neanche dire se esiste qualcosa di fattuale. Cioè, fatti e valori sono davvero *concettualmente* separabili, ma ogni insieme di fatti che io concepisco deve essere accompagnato dagli insiemi di valori che portano a scegliere quella concezione.

Ecco la rivoluzione copernicana: finora ho ragionato come se fossero i fatti a determinare i valori, ora invece sto dicendo che sono i valori a determinare i fatti.

Una conseguenza di questo ribaltamento è che, nel mio mondo, non esiste nulla di davvero neutrale. Quando guardo una bottiglia d'acqua, e vedo che è stata disegnata in un certo modo anziché in un certo altro, se ho un'adeguata sensibilità artistica capirò che quella scelta è resa possibile da certi valori. Quando guardo un uomo e vedo che ha scelto di farsi crescere i baffi invece di radersi, di tagliarsi i capelli in un certo modo, di comprare certi vestiti, intuisco anche che tipo di persona è, quali sono le cose che crede importanti, come vede il mondo. Qualsiasi scelta presuppone valori, dunque qualsiasi scelta, anche se di scarsa importanza, è espressione di una certa morale (che io potrei comunque fraintendere, visto che interpreterò il contesto tramite i valori miei propri).

Ulteriore conseguenza: è impossibile non giudicare, poiché tutti i concetti sono connessi a qualche valore (il che non significa che siano *semanticamente* spessi, il loro uso può rimanere del tutto descrittivo). Per questo, se io vedo una persona uccidere, penso e sento subito qualcosa di negativo in merito a quell'azione; d'altro canto, non sentissi qualcosa di negativo, avrei diversi valori, e avessi diversi valori la mia visione del mondo sarebbe diversa, perciò davanti a quell'azione potrei non vedere ciò che gli altri vedono.

Come dice Wittgenstein (2014), se un leone potesse parlare io non potrei capirlo, se lui mi parlasse di come ci si sente ad avere la coda io non lo capirei perché, non avendo accesso alla sua esperienza, io non posso esprimere dentro di me la proposizione che lui cerca di esprimere con le sue parole. Come dice Douglas Hofstadter (Hofstadter & Dennett 1992), dato che a lui non piace il cioccolato, non può neanche immaginare cosa voglia dire mangiarlo con piacere; e io non posso neanche immaginare cosa voglia dire giudicare bene un omicidio, in quanto non posso esprimere le

proposizioni che esprimerebbe chi sente così, o meglio potrei farlo solo se vivessi in modo *del tutto* diverso: citando sempre Wittgenstein, se una proposizione è una forma del linguaggio (cioè una delle cose da esso esprimibili), allora il linguaggio è una forma di vita, è vivere in un certo modo, cioè io potrei esprimere *esattamente* ciò che il leone vorrebbe dire solo se vivessi da leone.

La *sensibilità al contesto* sottolinea queste differenze d'interpretazione: usando un altro esempio di Wittgenstein (1981), quando guardo un oggetto bianco attraverso una lente rossa lo vedo rosso; similmente quando vedo un omicidio sulla strada, vedo appunto un omicidio e lo giudico male, ma quando lo vedo in un telefilm attraverso la lente del teleschermo, vedo un *omicidio** e può non essere male (perché potrebbe essere artisticamente lodevole); così, chi vede un omicidio sulla strada e non lo giudica male vede qualcosa di più simile a ciò che io vedo quando vedo un *omicidio**, che a ciò che vedo quando vedo un omicidio. Questo non vuol dire che il giudizio sia sbagliato: forse nel mondo ci sono davvero *omicidi** e non omicidi.

Dunque, per rispondere alla domanda iniziale, «stuprare a morte bambini per divertimento è male» non esprime una proposizione analitica, però la mia possibilità di concepire *stupro, morte, bambino* così come li concepisco *potrebbe* dipendere da una serie di valori che implicano la verità di quella frase. In altre parole, il concetto di «stuprare a morte bambini» non ha in sé il «male» che viene predicato, ma comunque la mia possibilità di possedere tale concetto potrebbe dipendere da valori che giudicano come male l'azione corrispondente. Quindi, in tal caso, un'azione di questo tipo può essere concepita solo come male, pensarla come bene è simile al sentire «gavagai» e capire *parti non-distaccate* invece di *coniglio*, o al percepire la mela come una parte del tavolo anziché un oggetto separato.

Dal punto di vista epistemologico, questa visione spinge verso l'intuizionismo: io ho un *sensu morale* che mi fa riconoscere subito il valore dell'azione, poiché se non lo riconoscessi non vedrei davvero *quella* azione. Questo, però, non è un metodo per ottenere conoscenza: io non posso conoscere l'etica indagando il mondo, al contrario posso conoscere il mondo perché conosco l'etica, e questo è il motivo per il quale esiste un disaccordo etico primitivo; d'altro canto, proprio perché gli uomini hanno una visione del mondo tutto sommato simile, anche i loro valori centrali devono essere simili, quindi non voglio certo dire che ogni uomo parla un linguaggio a parte in virtù di piccole differenze etiche.

Dal punto di vista ontologico invece, noto che i valori possono essere abdotti, ma non dedotti, dai fatti: da un lato, quindi, ha ragione il naturalismo perché, in primo luogo è vero che i valori, anche universali, sono inferibili da fatti naturali, e in secondo luogo ha ragione chi dice che ogni concetto è spesso in una certa misura; dall'altro, ha ragione il non-naturalismo, primo perché i valori precedono la conoscenza dei fatti naturali, e secondo perché i fatti naturali, proprio in virtù di certi valori che permettono di individuarli, potrebbero essere concettualizzati come moralmente neutri e non portatori di valori. Il mio concludere, in ultimo luogo, se i fatti etici siano naturali o non-naturali, anzi se vi siano fatti etici in primo luogo, dipende dalle risposte che dò alle altre domande della filosofia, dipende dal tipo di sistema che valori e fatti formano: ci sono valori che possono portare a credere a fatti che implicano la loro naturalità, e altri che portano ad implicare la loro non-naturalità.

I fatti vengono individuati a partire dai valori 2: conseguenze epistemologiche dell'etica

Questa particolare unità di fatti e valori non è di tipo ontologico, ma epistemico: non riguarda il modo in cui i fatti sono, e se ci sono valori tra di loro, ma solo il modo in cui posso conoscere tali fatti. Perché dovrei voler credere ciò che è vero, e non ciò che è falso? Perché fare previsioni scientifiche dovrebbe essere utile? Perché non credo in una realtà intrinsecamente imprevedibile? Perché voglio spiegare tutto, e non accetto un mondo inspiegabile? Perché voglio teorie coerenti e semplici, quando la realtà potrebbe anche essere incoerente e complessa? Queste domande sono epistemologiche, e la loro risposta è etica (cfr. Putnam 1985).

Che la conoscenza abbia anche un aspetto valoriale è un'intuizione che risale a Platone (2009 Meno), il quale si chiede perché la conoscenza abbia un maggior valore della credenza, perché dovrebbe scegliere di avere conoscenze anziché credenze vere. Putnam afferma che la conoscenza non dipende solo da fatti relativi alla credenza, ma anche nel suo valore, valore incarnato dalla giustificazione: per esempio, il coerentismo non si limita a notare il fatto che una certa credenza è coerente con le altre, ma esprime anche un giudizio, dà alla credenza coerente un valore maggiore, dà il dovere di favorirla rispetto a quelle incoerenti. Non c'è conoscenza senza teoria della giustificazione, e non c'è teoria della giustificazione senza giudizio di valore.

Ancora più in generale, qualsiasi ragionamento presuppone valori in quanto la logica stessa ha aspetti normativi. Esempio: l'abduzione. L'abduzione è valida se è una *inferenza alla miglior spiegazione*: un fenomeno come il fulmine può avere infinite spiegazioni, può essere causato da un evento fisico come anche dall'ira di Zeus, e dalla visione del fulmine io posso abduire la *miglior* spiegazione. Ma come faccio a dire quale spiegazione è *migliore*? Devo creare una scala di *valori*, devo specificare quali caratteristiche considero *importanti* e poi valutare le varie spiegazioni. I valori, allora, non sono la miglior spiegazione di alcun fenomeno osservabile, ma se non vi fossero valori non vi sarebbe neanche alcuna *miglior* spiegazione in primo luogo.

Insomma, se anche un ente perfetto credesse correttamente in tutti i fatti non-etici, ma non avesse alcuna conoscenza etica, né il *sensu morale* che forma e giustifica le credenze relative, non solo non potrebbe derivare da quei fatti non-etici tutte le conoscenze etiche, ma anche le sue credenze non-etiche non costituirebbero vera conoscenza.

Alla luce di ciò, dovrebbe risultare evidente l'importanza della *tradizione* al fine della conoscenza. Scrive Putnam (1985) che «noi possiamo sperare di produrre una concezione più razionale della razionalità, o una concezione migliore della moralità solo se operiamo all'interno della nostra tradizione». Lui, che ha vissuto come un ebreo pur non credendo perché non poteva permettere che il lavoro di Hitler fosse completato per mano degli ebrei stessi, lo sapeva bene. Io credo nell'esistenza di fatti, e nell'esistenza di proposizioni, e nell'esistenza di una relazione di corrispondenza tra proposizioni e fatti, perché ho certi valori che mi permettono di applicare questi miei schemi mentali al mondo; ma questi valori e questi schemi mentali non li ho creati io, vengono dalla mia tradizione. Chiedermi se il nucleo della mia tradizione sia vero o falso non ha alcun senso perché, se non avessi questa tradizione, non saprei neanche cosa dovrebbe essere la verità. Anche se la conoscenza è coerenza, anche se posso modificare certi miei valori senza danneggiare i miei concetti, esiste comunque un nucleo di credenze fondative che non può essere negato. Se uno vuole davvero rifiutare la sua tradizione, allora deve rifiutare anche il significato condiviso delle parole, la logica, la pratica scientifica, il modo in cui individua gli oggetti, deve crearsi un mondo autistico tutto suo; ma se invece vuole comunicare, confrontarsi, capire e far avanzare la comprensione del mondo, allora deve agire all'interno della tradizione, e il fatto che diverse persone, nate in diverse tradizioni, abbiano visioni del mondo diverse, non è affatto un problema da estirpare.

Tuttavia, questo non deve spingere all'antirealismo. Anzitutto, i fatti esistono, esiste una realtà indipendente dalla mente, e questa realtà potrebbe prevedere anche fatti etici. È vero che il mio riconoscere l'esistenza stessa dei fatti (e se ve ne siano di etici o meno) dipende dai miei valori, e quindi tutte le mie argomentazioni a difesa del realismo sono contingenti; ma questo non vuol dire che io possa credere tutto ciò che voglio: se ci sono fatti, l'etica può essere fattuale, e se non lo è, comunque esistono sistemi di valori che portano a concepire il mondo così com'è, valori che portano a compiere scelte in modo tale da creare una visione del mondo vera, e questi sono i veri valori. Epistemologicamente, in virtù di questa loro basicità, le credenze sui valori potrebbero essere gli unici veri posit (dato che anche la logica, che avevo proposto come posit, è derivata da esse), oppure potrebbero essere inferite da credenze su fatti naturali (alcune delle quali saranno a loro volta inferite dalle credenze sui valori); in ogni caso, io non posso credere *coerentemente* in una qualsiasi combinazione di fatti naturali ed etici, la credenza in certi fatti naturali richiama la credenza in certi valori e viceversa.

Essendovi unità tra fatti e valori, ed essendo la giustificazione basata appunto su valori, si può pensare che le due teorie normative, cioè teoria della giustificazione ed etica normativa, formino in realtà un'unica teoria. Non è proprio così, sono necessarie delle specificazioni.

I miei valori implicano sia la mia etica, sia la mia teoria della giustificazione. L'etica normativa è sia una teoria generale del bene, sia una teoria particolare della giustificazione: cioè alla domanda «è bene?» risponde in modo valido per tutte le azioni, mentre alla domanda «sono giustificato a crederlo?» risponde solo per le credenze relative alle azioni. Al contrario, la teoria della giustificazione alla domanda «sono giustificato a crederlo?» risponde in modo valido per tutte le credenze, mentre alla domanda «è bene?» risponde solo per le azioni relative alla credenza. In altre parole, la domanda generica «sono giustificato?» equivale in etica alla specifica «è bene che io lo creda?», mentre la domanda generica dell'etica normativa «è bene?» equivale alla specifica «sono giustificato a crederlo bene?» in teoria della giustificazione. Si può così notare, per esempio, che la domanda «è bene che io sia giustificato?» equivale all'assurda «è bene che sia bene che io lo creda?», e questo mi conferma che la giustificazione contiene già una dimensione etica.

Di conseguenza, non posso usare i miei criteri etici come generali criteri di giustificazione: per esempio, non posso dire che una credenza è giustificata se rispetta l'imperativo categorico, perché l'imperativo categorico non è truth-conducive per ogni credenza. Al massimo, potrei usare la mia etica normativa come *etica della credenza* per scegliere tra due credenze che sarebbero altrimenti ugualmente giustificate, cioè potrei usarla solo per dare un *surplus* di giustificazione.

Ciò nonostante, le due teorie devono lavorare a sistema: da un lato, il mio usare una certa teoria della giustificazione deve essere un bene; dall'altro, la mia credenza in una certa teoria etica deve essere giustificata. Dirò allora così: il fatto che sia bene credere x non significa che io sia giustificato a crederlo, e il fatto che io sia giustificato a credere che credere x è bene non significa che sia bene crederlo, ma se sono giustificato a crederlo, allora è bene che io lo creda, e se è bene che io lo creda, allora sono giustificato a credere che sia bene crederlo. In simboli, indicando con G il bene:

$$Jx \rightarrow Gx \rightarrow JGx$$

Immaginare giustificazione ed etica a sistema vuol dire anche introdurre elementi etici emotivi e quindi soggettivi nella conoscenza. Questo non è problematico. In primo luogo, la teoria della giustificazione accetta che ciò che giustifica un individuo può non giustificarne un altro, ciò che è conoscenza per uno può non esserlo per un altro: il fatto che un fisico premio Nobel creda nell'esistenza degli elettroni è sufficiente a giustificare me nel crederci, ma non può essere

sufficiente a giustificare altri fisici. In secondo luogo, come evidenzia Putnam (1985), la giustificazione è, in essenza, una *sensazione* che viene poi razionalizzata, e questa sensazione può essere intesa anche come un'emozione, l'emozione della giustificazione, e quindi qualcosa che ha già elementi soggettivi.

Si può obiettare che così si finisce per far traballare la distinzione tra oggettivo e soggettivo; ma anche questo non è un problema. Un albero che esiste e continuerebbe ad esistere anche se io non ci fossi è qualcosa di oggettivo? Fino ad un certo punto: se non ci fossi io coi miei valori, non ci sarebbe nessuno a *individuare* l'albero, ad applicare gli schemi mentali che distinguono l'albero dal terreno e dal cielo. Di contro, un fatto sociale come un matrimonio è soggettivo, esiste solo nella mente? Filosofi quali Searle (2006) e Maurizio Ferraris (2014) rispondono di no: il matrimonio non è solo un pensiero, è un evento con una estensione fisica e temporale, trascritto su documenti, che richiede la presenza fisica di certe persone, e il fatto che una società riconosca certe persone come sposate, le pensi come sposate, è un fatto oggettivo e assoluto, non è qualcosa il cui valore di verità cambia da persona a persona. L'albero, oggettivo e naturale, ha anche elementi soggettivi e sociali, mentre il matrimonio, soggettivo e sociale, ha anche elementi oggettivi e naturali: non è quindi uno scandalo che la conoscenza dei fatti possa dipendere da fattori sia oggettivi che soggettivi.

Pluralismo e soggettivismo non sono alternative al realismo (ma il relativismo è inconsistente)

Quando formulo un giudizio creo una funzione che collega l'azione al suo valore. È come se dentro di me avessi una macchina che incarna una certa teoria etica: io dò gli eventi in pasto alla macchina, lei li mastica, li elabora, e poi sputa un numero, che è il valore che dò a quell'azione.

Ad esempio, usando lo stesso stile visto per le teorie della verità, posso esprimere l'utilitarismo così:

- x è bene sse ha le conseguenze più felici.

La macchina così programmata osserva l'azione, ne immagina le conseguenze, e in base a quante sono e quanto felici calcola un numero più o meno grande, mi dice *quanto è bene* quell'azione

Discutendo la teoria della verità mi sono fatto una *metadomanda*, mi sono chiesto se la teoria sia a sua volta vera. Adesso devo rifarmi questa domanda in senso etico. Non posso chiedermi se il bene sia bene perché il bene è una proprietà delle azioni, non può essere proprietà di sé stesso; posso però chiedermi: il *credere* ad una certa teoria etica è a sua volta etico? Il credere che il bene sia l'averle le conseguenze più felici è a sua volta un bene? Una possibile risposta è questa:

- « x è bene sse ha le conseguenze più felici» è bene sse è un mio dovere

Questo vuol dire che la mia macchina sta usando un criterio deontologico per esprimere un giudizio sull'uso dell'utilitarismo. Se credo tale proposizione vera, credo che la deontologia giudichi il sostenere l'utilitarismo come un bene.

Vale però quanto già detto in epistemologia: questa proposizione è accettabile solo se i due «bene» sono logicamente equivalenti (e non lo sono, il dovere è indipendente dalle conseguenze) oppure se appartengono a metalinguaggi di livello diverso; in quest'ultimo caso, può essere che i due «bene» abbiano significati diversi, che il primo indichi il *bene* e il secondo un *bene**, ma allora non

parlo più di una teoria del bene, bensì di due teorie diverse, oltre all'etica dovrei sviluppare un'etica*. Quindi, quando esprimo una teoria del bene, affinché questa sia coerente e desiderabile le due parti devono essere uguali:

- «x è bene sse ha le conseguenze più felici» è bene sse ha le conseguenze più felici

La deontologia soddisfa senza problemi questo requisito in quanto pone sé stessa come dovere: infatti, se all'interno della deontologia io mi chiedo «è per me legittimo non seguire la deontologia?» ottengo una domanda assurda in quanto, essendo la deontologia a determinare ciò che devo fare, sarebbe come chiedere «è legittimo venire meno ai miei doveri?» L'utilitarismo può soddisfarlo solo in modo rocambolesco: se si accetta l'esternalismo della motivazione il mio *credere* la teoria non è sufficiente a causare la sua *attuazione*, perciò l'utilitarismo potrebbe riconoscere che è bene *attuare* l'utilitarismo ma che sarebbe meglio *credere* ad una teoria diversa, il che è assurdo; questo paradosso si può evitare accettando la possibilità di compiere un male se questo è presupposto (non causa) della possibilità di compiere un bene futuro (ma questo forse va già al di là dei principi dell'utilitarismo).

Sono invece da rifiutare tutti i sistemi che giudicano altri sistemi (e quindi danno ai due «bene» un senso diverso) ponendoli sullo stesso livello di metalinguaggio, un errore caratteristico del *relativismo*.

Cos'è un'etica relativista? È una teoria che accetta più teorie, che accetta vi siano più modi di assegnare valore ad una stessa scelta, cioè che accetta vi siano più *beni*: qualsiasi giudizio uno esprima, esso è relativo nel senso che esiste un altro sistema di pari valore che esprime un giudizio opposto. È male uccidere? È relativo: questo sistema dice di sì, ma un altro, giudicato da questo come altrettanto valido, dice di no. Il relativismo etico non è anche epistemologico, riguarda il valore e non la verità o falsità delle teorie, il relativista etico è colui che non può riconoscere la propria teoria etica come la teoria di più alto valore.

Ponendo più teorie sullo stesso livello di metalinguaggio, il relativismo finisce in un paradosso simile a quello dello Yeti visto nella teoria pluralista della verità: se ho due teorie, A e B, tali che A dice di tirare la leva nel problema del carrello ferroviario e B dice di non farlo, e A, essendo relativista, dice che agire in accordo con B è bene tanto quanto agire in accordo con A, la teoria A affermerebbe sia che bisogna tirare la leva, sia che (tramite B) non bisogna farlo: è come se si dicesse «se A è vero allora B è vero, e se B è vero allora A è falso».

Il relativismo, in più, è incapace di giustificare la credenza in esso. Una teoria etica, per essere relativista, come minimo deve valutare il credere in una teoria non-relativista quanto il credere a sé stessa: infatti, se presentasse solo teorie relativiste come di valore, e rifiutasse tutte le teorie non-relativiste, il relativismo non sarebbe a sua volta relativo, sarebbe un assoluto. Ora, la teoria relativista mi dice che posso sostenere anche una teoria non-relativista ugualmente buona; la teoria non-relativista, invece, si presenta come la miglior teoria possibile, e mi impedisce di sceglierne un'altra: perché dunque dovrei sostenere una teoria relativista, quando questa non giustifica lo sceglierla mentre tutte le altre lo impediscono?

Una teoria etica, per essere razionalmente accettabile, deve presentare sé stessa come la teoria di più alto valore; può accettare altre teorie come propria parte, teorie che essa implica in quanto perfettamente concorde con tutto ciò che affermano, ma qualsiasi teoria non-equivalente a sé o a proprie parti deve essere rifiutata. Questo rende il relativismo impossibile, ma non rifiuta la possibilità del *pluralismo*.

Il pluralismo trova in etica un terreno compatibile che lo rende più facilmente accettabile che in epistemologia: mentre una proposizione deve essere o vera o falsa, un'azione potrebbe non essere un bene ma neppure un male; mentre è impossibile che una proposizione e la sua negazione abbiano un uguale valore di verità, è possibile che esprimano comportamenti di uguale valore etico. Questo pluralismo etico è ben distinto dal relativismo: non accetta la validità di più teorie, afferma un'unica teoria con più valori; non accetta più modi di assegnare valore ad una stessa azione, afferma solo l'esistenza di più azioni di uguale valore; non accetta l'esistenza di più *beni*, c'è un unico bene che si realizza in virtù di fatti diversi.

Tutte le teorie etiche che ho esposto sono pluraliste in questo senso. Nell'utilitarismo è possibile che due azioni diverse abbiano conseguenze dello stesso valore, e siano quindi entrambe un bene. In deontologia ho immaginato contesti neutri nei quali né il compiere un'azione, né il non compierla, possono costituire doveri, di modo che entrambi i comportamenti siano legittimi. L'etica della virtù è pluralista per definizione: in uno stesso contesto diverse scelte possono istanziare diverse virtù, buttarsi tra le fiamme per salvare un cagnolino è eroico, non farlo è prudente, restare fuori a chiamare i pompieri è sagace; cioè diverse azioni sono rese tutte bene da diverse virtù (posso dire che hanno diversi *goodmaker*). Anche le etiche religiose sono spesso pluraliste: nel cattolicesimo occidentale sia costruirsi una famiglia, sia diventare sacerdote, sono buone scelte, ma una è incompatibile con l'altra (anche se non è vero per tutti: i cattolici maroniti possono fare entrambe le cose).

In casi come questi, davanti a due strade di uguale valore, per scegliere bisogna andare *oltre* l'etica, serve uno *slancio di volontà* (cfr. Schopenhauer 2011, Nietzsche 1886, Ciraci 2006, Vitale 2014). Questo inoltre è utile a spiegare i dilemmi e il disaccordo: è irrealistico pensare di essere sempre in grado di assegnare ad ogni possibilità d'azione un valore diverso e riuscire così ad individuare un'unica azione migliore di qualsiasi altra, la mia esperienza mi rende ovvio che esistono più opzioni di uguale valore, per questo ho esperienza del dubbio, e per questo persone ugualmente razionali possono giungere a giudizi diversi (cfr. Palmira 2014). Per quanto possa essere particolareggiato, un sistema etico di questo tipo non può essere sempre sufficiente al decidere come agire.

Al che, posso formulare una teoria pluralista d'esempio:

- x è bene sse [è coraggioso, o è magnanimo, o è giusto...].

Come in epistemologia, la definizione tra parentesi indica un unico bene, e non tipi diversi di valore. È come dire:

- x è bene sse è virtuoso

dove è virtuoso tutto ciò che è coraggioso, tutto ciò che è magnanimo, tutto ciò che è giusto, eccetera: ogni virtù ha una sua estensione comprendente diverse azioni, «bene» ha un'estensione composta da tutte queste estensioni, il bene è plurimo perché individuato tramite una serie di sottoinsiemi.

Il pluralismo permette di creare un'unica teoria fondendo più teorie in modo coerente. Per esempio, posso aggiungere più criteri alla lista delle virtù:

- x è bene sse [è virtuoso, o rispetta l'imperativo categorico, o ha le conseguenze più felici].

Ciò che si fa, in tal caso, è fondere l'estensione delle azioni riconosciute come buone dalle varie teorie alla super-estensione di «bene»; il principio «se devo allora posso» non crea problemi perché il suo ruolo è quello di determinare l'estensione della deontologia, e non quella di «bene».

Un'altra possibilità è quella di dividere i contesti in classi, di modo che in diversi casi la scelta dell'azione dipenda da teorie diverse. Per esempio, se accetto che l'etica kantiana fosse intesa solo per la morale pubblica, e non per le decisioni private (cfr. Varden 2010), posso immaginare un sistema pluralista di questo tipo:

- x è bene sse [(se è una decisione pubblica, se rispetta l'imperativo categorico) e (se è una decisione privata, se ha le conseguenze più felici)].

In tal caso si ottengono teorie coerenti perché si sommano le azioni riconosciute come bene da più teorie, ma si ignorano le azioni che le teorie presentano come male. Se io affermassi, per esempio, che «è virtuoso aderire alla deontologia» sosterei una teoria della virtù che mi chiede di seguire la deontologia, e in tal caso dovrei accettare tutti i giudizi, sia positivi che negativi, di entrambe le teorie; di conseguenza, una stessa azione potrebbe risultare sia bene che male, e ciò è inaccettabile. Questo non accade sommando solo i giudizi positivi e lasciando il discorso sui negativi a parte (verranno determinati poi per esclusione), il fatto che azioni opposte risultino ugualmente bene non è incoerente.

Il *soggettivismo* è un caso particolare del pluralismo, è quel pluralismo che include nei suoi criteri per l'individuazione dei sottoinsiemi dell'estensione di «bene» elementi soggettivi, cioè il pluralismo che include tra i fatti che possono rendere vere le proposizioni etiche anche fatti soggettivi (stati mentali, convenzioni sociali o altro). Per esempio, la *virtù*, in epoca classica, era sì l'insieme di qualità desiderabili, ma principalmente *per un uomo*, l'equivalente più comune per la donna era la *pudicizia*, definire una donna virtuosa era possibile ma raro quanto definire un uomo pudico. Al che, posso rappresentare quest'etica soggettivista così:

- l'azione x compiuta dall'agente y è bene sse [(se y è un uomo, se è un'azione virtuosa) o (se y è una donna, se è un'azione pudica)]

Nelle teorie normative che ho esposto ci sono anche elementi di soggettivismo. Nell'etica della virtù ogni agente deve agire in modo personale, non deve imitare il virtuoso ma bensì considerare le particolarità della sua condizione. Nella deontologia ho ipotizzato l'esistenza di doveri *ad personam* che dipendono da condizioni personali, una madre ha verso suo figlio doveri speciali che non ha verso altri né altri hanno verso di lui. L'utilitarismo cerca di essere oggettivo e imparziale e quindi senza elementi soggettivi, ma comunque può riconoscere che, se io sono un bravo matematico ma un pessimo carpentiere, è meglio che stia a fare i conti invece di costruire un muro, perché così potrò lavorare meglio e creare un futuro positivo, cioè il valore delle conseguenze dipende anche da elementi personali. Nell'etica religiosa pure ci sono elementi soggettivi: per la morale cristiana uccidere è peccato mortale, ma se chi uccide è incosciente, o lo fa per legittima difesa, allora non è più così; similmente, scegliere se diventare sacerdote o padre di famiglia dipende da un fattore personale, la *vocazione*, da individuare tramite *discernimento*; addirittura, se il bene dipende dalla volontà di Dio, essendo la volontà un fattore personale, tutta l'etica religiosa è soggettiva (pur essendo universale).

Il soggettivismo, per assurdo, deve mantenere oggettività: è coerente perché si esprime tramite condizionali, non dà mai valori opposti ad una stessa azione *simpliciter*, dice bensì che se si dà una certa condizione ha un certo valore, se si dà un'altra condizione (che non può darsi assieme alla

prima) ha un altro valore, ed è oggettivo che abbia un certo valore in una certa condizione. Ciò nonostante, un soggettivismo *totale*, cioè un'etica in cui il valore di tutte le azioni cambia in base a fatti soggettivi, è inaccettabile: se ci sono valori universali, questi devono essere *transculturali* e *transpersonali*, quindi indipendenti da fattori personali, certe azioni devono essere associate ad uno stesso valore in qualsiasi condizione.

LETTURE PER APPROFONDIRE

Un intuizionista vicino alle mie posizioni è W. D. Ross: si vedano 1930 e 1939. Ross è un pluralista, ma il più famoso tra questi (anche se per il pluralismo politico, più che metaetico) è il contro-illuminista Isaiah Berlin: tra i suoi molti libri si vedano 1959, 1994, 1997, 1998, 2001.

Per approfondire aspetti secondari di alcune argomentazioni viste: per i concetti spessi, Williams 1995; per la direzione di adattamento, Searle & Vanderveken 1985 e Searle 1985; per la motivazione, Smith 1987; per la supervenienza, Hare 1982 e 1993; per la Terra* gemella, Sayre-McCord 1997; per la distinzione tra oggettivo e soggettivo, Searle 2010.

Introduzione ai ragionamenti di metafisica

Quali argomenti affronterò, e quali no

L'altisonante nome «metafisica» ha origini umili: significa «oltre la fisica» in quanto gli argomenti caratteristici di questo settore sono quelli del libro di Aristotele che il curatore della sua opera, Andronico di Rodi, ha posto dopo il libro «Fisica». Aristotele, dal canto suo, chiamava questi argomenti «filosofia prima», non «metafisica» (Cohen & Reeve 2020).

È un cattivo nome, e un cattivo posizionamento per il libro: doveva essere *prima*, non *dopo* la fisica. La metafisica non si occupa di ciò che è *oltre il fisico* od *oltre il naturale* come può fare la teologia. Non è neanche un discorso *di secondo livello* o un *ragionamento sul ragionamento* come il prefisso «meta» potrebbe far pensare (come la *metalogica* che è l'uso della logica per analizzare la logica). La metafisica è il settore che si occupa delle domande più generali e astratte possibili, quelle che riguardano l'*esistenza* in sé e nel suo complesso. Cioè, come dice il frate domenicano Tomáš Týn (1986), la metafisica è quella materia che ragiona sull'universo tutto come se fosse un unico oggetto, che ragiona sulla realtà in generale e non su elementi isolati. La metafisica, quindi, deve precedere la fisica nel senso che si occupa di quei concetti generalissimi che fanno da fondamento a tutte le altre imprese intellettuali: la metafisica discute ciò che tutte le altre materie presuppongono.

Nella prima metà del Novecento la metafisica era malvista dalla tradizione analitica. Come dice Hume (1748) con la sua *forchetta*: «se ci viene alle mani qualche volume, per esempio di teologia o di metafisica scolastica, domandiamoci: contiene qualche ragionamento astratto sulla quantità o sui numeri? No. Contiene qualche ragionamento sperimentale su questioni di fatto e di esistenza? No. E allora, gettiamolo nel fuoco». Rasoio di Occam in una mano e forchetta di Hume nell'altra, i positivisti logici hanno fatto a fette un intero settore, rifiutando ogni ragionamento metafisico (Carnap 1996).

Oggi le cose non potrebbero essere più diverse: la metafisica è ora il settore più ampio, complesso e frizzante di tutta la filosofia analitica, quello in cui tutti gli studenti desiderano specializzarsi, e nessuna tradizione continentale ha discusso i tradizionali problemi metafisici altrettanto a lungo e approfonditamente (cfr. Bourget & Chalmers 2014). Grazie soprattutto al lavoro di David Kellogg Lewis, che ha dato alla disciplina un notevole rigore matematico, la metafisica si è guadagnata un così grande rispetto da essere considerata da alcuni *come la fisica, ma più profonda*, e questo è proprio ciò che anche io penso.

Dunque, di cosa parla questa metafisica?

L'*ontologia*, cioè lo studio dei concetti di *essere*, *esistenza*, *realtà*, è la storica corona della metafisica. Nella tradizione analitica questi temi non sono affrontati direttamente, sono bensì discussi tramite gli altri settori. Oggi si ragiona soprattutto su questo tema: se le domande sull'*essere* siano problemi filosofici complessi (tale è l'idea del *massimalismo*), o se non siano altro che problematiche grammaticali destinate a scomparire una volta analizzato a fondo il verbo «essere» (*deflazionismo*) (cfr. Chalmers et al. 2009).

Uno dei principali compiti dell'ontologia è lo sviluppo di sistemi di *categorie*: Linneo ha classificato gli esseri viventi secondo specie, genere, famiglia, ordine, classe, e via così verso categorie sempre

più generali; similmente fa l'ontologia ma puntando alla massima generalità, punta a sviluppare la tassonomia di tutto ciò che esiste, tutto ciò che è pensabile, tutto ciò che è dicibile su qualcosa, sia esso concreto o astratto. Aristotele (2003) e Kant (1781) hanno sviluppato influenti sistemi di categorie, l'ultimo sistema rilevante lo si deve a Peirce (2011 §1.545-59). Oggi la creazione di sistemi di categorie costituisce un importante settore dell'informatica: la rappresentazione e organizzazione della conoscenza è infatti fondamentale, tra le altre cose, allo sviluppo di intelligenze artificiali in grado di classificare gli enti del mondo in modo simile a quello umano (cfr. Smith 1982).

Altro compito dell'ontologia è l'analisi delle proprietà (le caratteristiche, i modi d'essere degli enti), e determinare se vi siano proprietà *essenziali*, cioè le proprietà che un certo ente non può avere perché senza di esse non sarebbe più ciò che è, oppure (come le definisce Gödel 2006) proprietà che implicano tutte le altre proprietà di quell'ente.

Di particolare importanza sono anche le proprietà *universali* (opposte alle *particolari*). Un universale è una caratteristica comune a più oggetti: per esempio, il *rosso* è un universale perché comune a tutti gli oggetti rossi (Aristotele 2003). La domanda è: *cosa* stanno condividendo questi oggetti? Secondo i *realisti* come Platone, gli universali sono entità la cui esistenza è indipendente dagli oggetti particolari, anche sparissero tutti gli oggetti rossi il *rosso* continuerebbe ad esistere in un diverso livello della realtà; secondo i *nominalisti* come Occam e il monaco Roscellino, gli universali sono categorizzazioni e astrazioni dell'immaginazione, esistono gli oggetti rossi ma non il *rosso*; secondo i *moderati* come Aristotele, gli universali esistono realmente ma in modo *dipendente* dagli oggetti, il *rosso* esiste ma solo finché esistono oggetti rossi (Klima 2017).

Un altro settore della metafisica è quello della *cosmologia*. La cosmologia filosofica si occupava proprio di ciò che sembra: tentava di descrivere la struttura dell'universo, la distribuzione della materia nel cosmo. La cosmologia intesa come scienza non esiste fino al Settecento, è noto che anche Newton definiva sé stesso «filosofo della natura» e non «scienziato». Anche se oggi l'indagine cosmologica, divenuta del tutto empirica, si è separata dalla metafisica, negli ultimi anni ha preso piede una nuova *filosofia della natura* che si interroga su tre temi principali: quale sia l'origine dell'universo (se abbia senso parlare di una *causa prima*, di *causa incausata*, o se sia invece eterno), quale sia il suo costituente più fondamentale (la materia, l'energia, i campi, le *monadi*...), e se la sua esistenza abbia uno scopo (cfr. anche Smeenk & Ellis 2017).

L'analisi dei concetti di *possibilità* e *necessità* costituisce un altro importante settore metafisico. La filosofia analitica affronta questo campo tramite la *logica modale aleatica*, una formalizzazione matematica di questi concetti. Nonostante le iniziali ostilità di Russell e Quine verso la logica modale (Rescher 1979, Quine 1986), i lavori di Kripke (1999), David Lewis (2020), Plantinga (1974) e altri l'hanno sviluppata al punto da renderla un mezzo indispensabile alla metafisica analitica.

L'interpretazione standard della logica modale usa il concetto dei *mondi possibili*: se lancio un dado, esiste un mondo possibile in cui esce 1, uno in cui esce 2, eccetera (il termine è di Leibniz, e.g. 1710). Se il dado esiste in tutti i mondi possibili dirò che esiste *necessariamente*, se vi sono mondi senza quel dado dirò che è solo *possibile*, se non vi sono proprio mondi col dado allora è *impossibile*. L'importanza di questi concetti è evidente lungo tutta la storia della filosofia: il grande teologo Anselmo d'Aosta (2002) ha creato la sua prova dell'esistenza di Dio partendo dall'idea che un ente concettualizzato come perfetto debba essere dotato di esistenza *necessaria*; quasi mille anni dopo Chalmers (1996), influente filosofo contemporaneo, formula la sua più importante

argomentazione sulla distinzione di mente e corpo basandosi sulla *possibilità* di corpi viventi ma privi di mente.

Ma questa logica modale è più ampia di così, non è certo limitata alla discussione su *possibilità* e *necessità*: Hughes e Cresswell (1996) individuano ben 77 tipi diversi di logica modale, e sicuramente oggi ce ne sono ancora di più, tanto che si potrebbe parlare di un intero settore della *metafisica modale*. Ci sono infatti logiche modali che analizzano i concetti di *obbligo* e *permesso* (logica *deontica*, e.g. Mally, si veda Lokhorst 2019), di *credenza* (logica *dossastica*, e.g. Smullyan 1986), di *conoscenza* (logica *epistemica*, e.g. Rescher 2005), di *passato*, *presente*, *sempre*, *a volte* (logica *temporale*, e.g. Prior 2003). Oggi la logica modale ha molte applicazioni anche in informatica e nelle scienze umane (e.g. Pnueli 1977).

Un altro tema capitale della metafisica è l'analisi del concetto di *causa*. La prima trattazione in merito risale ad Aristotele (2010), con le sue celebri *quattro cause*: un tavolo ha una *causa materiale* (i suoi atomi, il suo legno), una *causa formale* (la sua struttura molecolare, l'essere un piano con quattro gambe), una *causa efficiente* (l'essere creato dal lavoro del falegname), e una *causa finale* (il puntare ad essere usato per mangiare). È chiaro che Aristotele non parla di *causa* nel senso moderno, ciò che intende è in realtà *spiegazione*, le quattro cause sono quattro spiegazioni al perché il tavolo è così com'è, sono quattro modi di rispondere ai «perché?» e quindi quattro cose che devo sapere per capire il mondo e i suoi cambiamenti (Falcon 2019). Questa concezione della causalità domina il pensiero fino al Rinascimento, solo in seguito si sviluppa il concetto moderno di causa intesa come il modo in cui il mondo evolve e cambia, il modo in cui gli eventi passati pongono limiti ai possibili eventi futuri.

Ci sono oggi tre principali teorie della causalità, che non si escludono vicendevolmente. La prima è la teoria *controfattuale*: l'evento x causa l'evento y nel senso che, se non si fosse dato x , non si sarebbe dato y (ma, se si dà il primo, deve darsi il secondo); il lavoro del falegname *causa* il tavolo perché senza il primo non ci sarebbe stato il secondo. In altre parole, x causa y se x è necessario a y (cfr. Lewis 1973).

La concezione controfattuale di causa è molto ristretta, non rende merito di tutti gli usi della parola: il fumo causa il cancro, ma ci sono anche pazienti oncologici che non hanno mai fumato, e altri che fumano senza mai sviluppare un simile male, quindi il fumo non è necessario come richiederebbe la teoria controfattuale. La teoria che rende merito di questo uso è quella *probabilistica*: il fumo causa il cancro nel senso che conosco il meccanismo per il quale fumare aumenta la probabilità di sviluppare il cancro in futuro. Oggi i rapporti di causazione così intesi sono formalizzati matematicamente e analizzati a computer tramite la statistica bayesiana, con tecniche che rendono questo settore molto tecnico, interdisciplinare, e di interesse scientifico (Neapolitan 2004).

La terza teoria è quella *humeana* che ho già esposto: ciò che chiamo causa è solo una *ricorrenza*, l'evento x causa l'evento y solo nel senso che *di solito* ad x segue y , *di solito* al calcio segue il volo del pallone, ma non è affatto detto che in futuro sarà così (Hume 1739).

Bisogna fare attenzione a non fraintendere Hume. Riprendo il *problema dell'induzione*, al quale ho già accennato tramite il falsificazionismo di Popper: se io osservo un evento e vedo che ne segue un altro, e ripeto questa osservazione più volte, induco che il primo causa il secondo, ma il fatto che io finora abbia visto solo cigni bianchi non vuol dire che non ne vedrò mai di neri; cioè, come diceva Einstein (2002), mentre nessun insieme di prove può dimostrare vera una teoria, ne basta solo una per dimostrarla falsa, e quindi nessuna serie di osservazione può giustificarmi nel credere

nell'esistenza di un rapporto di causazione tra due eventi. Ma l'uso dell'induzione per individuare rapporti di causazione è la base della scienza naturale, se si nega che tale operazione possa condurre verso la verità tutta l'impresa scientifica crolla. Come salvarla?

Hume, che era uno scettico ma anche un naturalista, non desiderava negare la causazione, così, come nota Kripke (2000), individua una *soluzione scettica* che rende legittima l'induzione del rapporto di causa a prescindere dai fatti: se io prendo due eventi singoli, posso dire solo che ad uno segue l'altro, non posso dire che uno causi l'altro, primo perché non c'è alcun fatto che renda la successione una necessità (facendola così divenire causazione), secondo perché anche se tale fatto vi fosse non potrei conoscerlo; ma se prendo invece classi di eventi, e vedo che ad ogni evento elemento dell'insieme A segue uno dell'insieme B , allora esiste una funzione che assegna ad ogni *evento-A* un *evento-B*, e questa funzione costituisce la causazione.

Come ultima nota, David Lewis, famoso per rappresentare i rapporti di causa tramite schemi *a neuroni* (che non sono altro che pallini collegati da frecce), ha dedicato parte dei suoi ragionamenti alla relazione tra causazione e temporalità, e si è chiesto se una causa possa mai essere temporalmente successiva al suo effetto, cioè se un evento futuro possa *retrocausare* un evento passato o presente (Lewis 1976). Nonostante l'idea paia assurda e impossibile tanto da poter porre la sua falsità come assioma, negli ultimi decenni la filosofia analitica è arrivata alla conclusione che non c'è alcunché di paradossale nel concetto di retrocausazione, e anzi è un'interpretazione possibile delle osservazioni e delle teorie scientifiche (Dummett & Flew 1954).

Di tutti questi argomenti così stimolanti non parlerò più. Ho scelto di trattare solo alcuni dei temi della metafisica, e già così questa sezione occuperà più pagine di tutto il resto del libro. Ho prediletto i settori più attivi e discussi, quelli nei quali è più evidente il conflitto tra posizioni alternative: credo sia divertente confrontare le varie scuole, fonderle, prendere una posizione, come ho fatto finora. Nello specifico, mi dedicherò a questi quattro temi:

- filosofia dell'identità e del tempo;
- filosofia del libero arbitrio;
- filosofia della matematica;
- filosofia della mente.

Identità, spaziotempo e libero arbitrio sono temi d'importanza generale che presentano connessioni con tutti gli altri argomenti sia metafisici che etici ed epistemologici. La filosofia matematica è un settore molto tecnico, generalmente discusso da matematici di professione, che affronterò a grandi linee; è però molto importante perché al suo interno vi si trovano tutte le più importanti problematiche ontologiche, tanto da costituire un riassunto di tutta la storia della metafisica. Infine, la filosofia della mente è, al momento, il settore della filosofia analitica più ampio, complesso e praticato (cfr. Bourget & Chalmers 2014).

LETTURE PER APPROFONDIRE

I lavori dei primi scienziati possono essere considerati a cavallo tra cosmologia scientifica e filosofica: si vedano Keplero 1596 e 1619, Galilei 1632, Telesio 2009, Copernico 1979.

Sulla logica modale, si vedano anche Lewis & Langford 1932, Hintikka 1975, Plantinga 2001.

Sulla causalità, invece, si aggiungano Lewis 2000, Cartwright 2004, e Pearl 2000. Quest'ultimo è un personaggio interessante: informatico di professione, è il padre di Daniel Pearl, giornalista ucciso

da terroristi al quale Steve Reich, uno dei maggiori compositori viventi, ha dedicato uno dei suoi pezzi più godibili (Reich 2006).

Infine, per un esempio su come si pratici la metafisica analitica oggi, si veda Zalta 1983.

Metafisica: filosofia dell'identità e del tempo

Quinto ragionamento: un fusionismo minimalista

In questo capitolo propongo una teoria personale dell'identità che credo essere la base di tutto il mio pensiero; in quanto al tempo, invece, mi limito a tirare le somme pesando i pro e i contro. Il ragionamento è il seguente:

- P1** posso interpretare una totalità sia come un insieme di parti (cioè un insieme di enti che non formano un altro ente), sia come un unico ente divisibile in parti (che non sono enti);
- P2** la totalità non è convenzionale né relativa, e io sono una totalità;
- P3** se perdo una mia parte, io divento la totalità meno quella parte (e se aggiungo una parte, la totalità più quella parte);
- C1** quindi io non sono alcuno specifico insieme di parti, quindi vado inteso come un singolo ente divisibile in parti e non un insieme di parti (fusionismo);
- P4** né le proprietà corporali né quelle psicologiche sono essenziali, cioè entrambe possono cambiare senza che io perda la mia identità;
- P5** neanche la mia origine è essenziale;
- P6** io sono composto solo da proprietà corporali e psicologiche, con una loro origine;
- C2** quindi nessuna proprietà in me è essenziale, non c'è limite al mio cambiamento (minimalismo);
- P7** l'eternalismo non rende merito di tutti i fatti;
- P8** il presentismo non rende merito di tutti i fatti;
- C3** quindi è necessaria una diversa teoria del tempo.

Le due tesi dell'identità: endurantismo e perdurantismo

La persona che sta scrivendo questa riga è la stessa che ha scritto le righe precedenti, sono sempre io, sono sempre *lo stesso*; cioè, tra me e la persona che ha scritto ieri c'è una relazione di *identità*. Posso analizzare questo concetto così:

- l'ente *A* è identico all'ente *B* sse (se *x* è una proprietà di *A*, allora è una proprietà di *B*).

Cioè due enti sono identici se hanno esattamente le stesse proprietà: questa è la *legge di Leibniz*. Provo ora ad utilizzare il concetto di identità in diversi contesti:

1. $2+2$ è identico a 4;
2. questo elettrone è identico a qualsiasi altro elettrone;
3. io oggi sono identico a ciò che ero ieri.

L'analisi di Leibniz si adatta bene alla prima frase: il « $2+2$ » è inseparabile dal «4», sono identici perché, avendo tutte le proprietà in comune, risultano essere due rappresentazioni dello stesso concetto (si parla, in tal caso, di identità *numerica*). È un'analisi accettabile anche nel caso della seconda frase, che è invece una identità *qualitativa*: dico che due elettroni sono identici (o meglio, *identici**) nel senso che sono particelle dello stesso tipo, che condividono un certo sottoinsieme di proprietà (cariche, massa, spin) ma non tutte, le diverse proprietà spaziali o temporali (non possono occupare uno stesso punto dello spazio, in uno stesso istante) sono sufficienti a riconoscere che si tratta di due enti diversi, qualitativamente *identici** ma numericamente diversi.

Nel caso della terza frase, però, l'analisi non funziona. Si tratta di un'identità *transtemporale*, identità lungo il tempo. Io non ho le stesse identiche proprietà che avevo ieri, i miei capelli sono più lunghi per esempio, e di sicuro non ho le stesse proprietà che avevo venti anni fa, avevo un corpo diverso, diverse cellule, un cervello diversamente strutturato; eppure sono *la stessa persona*, sono numericamente identico a me stesso. L'albero che vedo dalla finestra ha le foglie, in inverno non le aveva, quindi *l'albero-in-estate* e *l'albero-in-inverno* hanno proprietà diverse; eppure sono lo stesso albero, c'è solo un albero lì fuori.

Leibniz cattura l'idea di identità come relazione che ogni ente ha solo con sé stesso e nessun altro (una concezione che Russell guarda con sospetto, ritenendola inutile alla metafisica). Questa però non può necessitare l'averne in comune *tutte* le proprietà: io sono identico solo a me stesso, ma continuo ad esserlo, continuo ad essere io, anche se perdo le gambe e trasformo il mio corpo e la mia mente. Posso provare a negare ciò, posso dire che in realtà non c'è identità, io non sono *io*, domani sarò una persona diversa; ma questo sarebbe contrario al senso comune. Quindi?

La soluzione sta nel distinguere (come fa uno dei maggiori metafisici del secolo scorso, Chisholm) l'identità *stretta* dalla *vaga*: « $2+2$ » è *strettamente identico* a «4» perché rispetta la legge di Leibniz, io invece sono ora *vagamente* (ma numericamente) *identico* all'io di ieri in quanto ho alcune ma non tutte le proprietà in comune. La filosofia analitica offre così due principali teorie dell'identità transtemporale, una legata all'interpretazione stretta e una legata all'interpretazione vaga.

La prima teoria è quella dell'*endurantismo*: ogni ente è pienamente presente, con tutte le proprietà che lo definiscono, in ogni istante del tempo in cui esso esiste; così, qualsiasi sia la proprietà che definisce l'albero fuori dalla mia finestra, esso continua ad esistere perché tale proprietà era presente lo scorso inverno, è presente questa estate, sarà presente il prossimo autunno. Basandosi su un sottoinsieme di proprietà, l'identità transtemporale risulta essere un caso di identità numerica ma vaga: l'ente cambia tanto che posso distinguere varie fasi della sua vita, ma quelle che cambiano sono solo proprietà *secondarie*, le sue proprietà *essenziali* continuano ad essere presenti in ogni istante nel quale quell'ente esiste.

La seconda teoria è quella del *perdurantismo*: l'ente che io indico con una certa parola esiste nel passato, esiste nel presente, ed esiste nel futuro, in quanto ogni istante del tempo contiene una sua *parte temporale*. In tale teoria, l'albero che vedo non è *tutto* l'albero, il vero albero è un ente esteso nel tempo, è un serpentone quadridimensionale del quale io posso vedere solo la sezione

tridimensionale intersecata dal sottilissimo piano del presente. Si può parlare, in questo caso, dell'identità transtemporale come numerica e stretta: l'albero è definito dal suo avere certe proprietà in inverno e certe altre in estate, è definito dal suo stesso cambiamento che è quindi illusorio, dovuto al fatto che non posso osservare gli enti nella loro atemporale interezza.

Le due tesi del tempo: presentismo ed eternalismo

Il problema dell'identità transtemporale è biforcuto: la sua soluzione dipende non solo dalla comprensione del concetto d'identità, ma anche dalla concezione del tempo adottata.

In passato lo spazio e il tempo erano concepiti come *assoluti*, cioè la loro esistenza e le loro proprietà erano intese come indipendenti da quelle di qualsiasi altro ente. Vi erano poche voci fuori dal coro. Galileo, per esempio, ha descritto dei casi di *relatività epistemica*: se io faccio un salto all'interno di una carrozza in corsa, mi pare di saltare sul posto, ma chi è fuori vede che compio un gran salto in lungo, salto sul posto relativamente ai passeggeri e in lungo relativamente agli spettatori esterni. Anche Leibniz ha offerto vari esempi a favore di una concezione *relativa* dello spazio, per lui il modo in cui posso descrivere il posizionamento e il moto di un oggetto dipende da quali altri oggetti ci sono. Ciò nonostante, il successo delle teorie di Newton, che sposava una concezione assoluta dello spazio, ha portato a ignorare queste interpretazioni per secoli, tanto che anche Kant considerava lo spazio e il tempo come le prime intuizioni dell'essere umano, le idee più basiche possibili. Con la teoria della relatività di Einstein le cose sono cambiate.

Io so che, se salissi su un'astronave in grado di viaggiare a velocità vicine a quella della luce, e vi rimanessi per qualche anno, al mio ritorno sulla Terra scoprirei che tutti i miei amici sono morti perché in realtà è passato un secolo: il *mio* presente non è più quello degli altri uomini, il *mio* tempo è passato molto più lentamente. Io so che, se precipitassi in un buco nero (e la gravità non mi distruggesse all'istante), mi limiterei ad avanzare nel cosmo, ma chi mi osserva dall'esterno mi vedrebbe invece rallentare fino a bloccarmi sull'*orizzonte degli eventi*. E poiché non c'è particolare differenza tra spazio e tempo, che sono solo diverse dimensioni di uno *spaziotempo* quadridimensionale, posso trovare anche analoghi spaziali di questi esempi: man mano che cado nel buco nero e la gravità aumenta, lo spazio attorno a me si schiaccia fino a diventare un foglio bidimensionale; ma chi mi osserva dall'esterno non vedrà nulla di tutto ciò, vedrà uno spazio sempre uguale. Che queste cose siano grossomodo corrette lo so perché lo si è visto negli esperimenti: orologi atomici precisissimi posizionati su aerei ad alta velocità perdono delle sia pur minuscole frazioni di secondo rispetto a quelli che rimangono fermi al suolo.

La conclusione che posso trarre da queste osservazioni è che né lo spazio né il tempo hanno una natura assoluta, sono solo relazioni tra gli enti, dipendono da quanto velocemente un ente si muove rispetto agli altri o da quanto è forte la sua gravità.

Voglio esporre un esempio che mi aiuti a capire la natura *relazionale* dello spaziotempo. Penrose, grande fisico e *mistico* contemporaneo, racconta un sogno. Tra miliardi e miliardi di anni tutta la materia verrà inghiottita da buchi neri, e poi, miliardi di anni ancora più tardi, tutti i buchi neri evaporeranno, lasciando nell'universo vuoto sola radiazione. Ma se non c'è più materia, non ci sono neanche relazioni tra particelle, quindi non c'è neanche lo spaziotempo, tutta l'energia è concentrata in un punto infinitamente piccolo al di fuori del tempo: da qui *fluttuazioni* casuali dei livelli d'energia porteranno ad un nuovo *Big Bang*, alla nascita di un nuovo universo e di un nuovo tempo. Vi sarebbe molto da specificare (non posso parlare qui delle definizioni di *particella*, dello

statuto ontologico delle *particelle virtuali*, di cosa siano i *campi*), ma il punto fondamentale è questo: con uno spaziotempo relativo se spariscono le relazioni sparisce anche lo spaziotempo.

Epperò, se è un errore credere che la discussione sulle dimensioni spaziali e temporali sia solo filosofica, lo è anche crederla solo fisica. Sia filosofi che fisici sono concordi nel dire che la verità di una teoria fisica è insufficiente a trarre conclusioni sulla natura dello spazio e del tempo: la teoria della relatività dice come essi *appaiono* in varie circostanze, dice che è *utile* considerarli relativi per fare previsioni, ma non può dire se siano enti o relazioni, se abbiano esistenza autonoma o dipendente, eccetera.

Il dibattito filosofico sulla natura del tempo inizia con McTaggart, un raro idealista rilevante anche per la filosofia analitica. A lui si deve la distinzione tra *serie A* e *serie B*, due diversi modi di parlare del tempo: in *serie A* passato e futuro sono assoluti e oggettivi, così posso dire «sono caduto *ieri*»; in *serie B*, invece, passato e futuro sono relativi, richiedono quindi di porre un punto di riferimento *rispetto al quale* l'evento di cui si parla è passato o futuro, così dirò «sono caduto un giorno *prima* di quando ho pronunciato questa frase».

La *serie A* si traduce nella *teoria A*: gli eventi seguono un ordinamento temporale oggettivo che va dal passato al futuro, gli enti sono in continuo cambiamento ma il tempo continuerebbe ad esistere e a scorrere anche se nulla cambiasse. La *serie B* si traduce invece nella *teoria B*: che un certo evento sia passato o futuro dipende da qual è il punto di riferimento, non c'è un ordine oggettivo degli eventi e quindi non c'è neanche alcunché che scorra da un passato verso un futuro, il cambiamento del enti è illusorio.

Serie e *teoria A* puntano al *presentismo*: com'è intuitivo, il passato è esistito, il futuro esisterà, ma solo il presente *esiste* in senso pieno. *Serie* e *teoria B* puntano invece all'*eternalismo*: passato, presente e futuro sono tutti ugualmente esistenti, ogni istante è come un diverso paese su una carta geografica, il tempo sembra scorrere ma tutto esiste già adesso alle proprie coordinate temporali.

Serie A e *serie B* riguardano il modo in cui si parla del tempo; *teoria A* e *teoria B* riguardano la struttura del tempo, se presenti un ordinamento oppure no; *presentismo* ed *eternalismo* riguardano lo statuto ontologico degli eventi passati e futuri. Non è logicamente incoerente pensare di combinare queste posizioni in modo poco ortodosso: posso essere presentista ma credere che si debba parlare usando lo stile della *serie B*, per esempio. Queste combinazioni sono però rare, quindi per praticità tratterò il presentismo come implicante *serie* e *teoria A*, e l'eternalismo come implicante *serie* e *teoria B*.

Essendo due facce dello stesso problema filosofico, c'è un'ovvia assonanza tra le teorie dell'identità e quelle del tempo: l'endurantismo rende merito dell'identità in un universo in continuo cambiamento, un universo in cui il presente cessa continuamente di esistere diventando passato e il futuro inizia ad esistere diventando presente, e per questo è una teoria che va a braccetto con il presentismo; il perdurantismo invece richiede l'esistenza di mie parti temporali nel passato e nel futuro, parti che rendono il mio cambiamento illusorio, e per questo va a braccetto con l'eternalismo. In ciò che segue, mi concentrerò prima sull'identità, poi tornerò a parlare del tempo.

La totalità non è convenzionale né relativa: Teseo e le sorite

Uno dei libri più stravaganti che ho è *S* di Doug Dorst (ma ideato dal regista J. J. Abrams). Il libro che fisicamente possiedo non riporta questi dati, ha invece una copertina con titolo *La Nave di Teseo* e come autore Straka, che non esiste. All'interno, le pagine sono tutte incorniciate da note scritte in più calligrafie, e tra le pagine si trovano oggetti come cartoline, fazzoletti, ritagli di giornale, santini. Tutto il corpo del testo è quello dello pseudobiblion *La Nave di Teseo*, che assieme a note e oggetti costituisce il libro reale *S*. È divertente perché fa riflettere sull'esistenza degli oggetti, su dove stia il confine tra un libro reale e uno solo immaginato.

Il titolo dello pseudobiblion fa riferimento ad un noto problema filosofico. Teseo ha una nave, si danneggia e bisogna sostituire un paio di assi di legno; passa del tempo, altre assi vanno sostituite, e poi altre ancora. Alla fine, tutte le assi sono state sostituite. La nave risultante è ancora la nave di Teseo, oppure, visto che tutto ciò che la costituiva è cambiato, è diventata un'altra nave? Ora prendo le assi vecchie e le riordino per formare un'altra nave: qual è la nave di Teseo? Quella con le assi nuove, o quella fatta con tutte assi originali?

La nave di Teseo si collega al *problema delle sorite* (in greco *σωρός* vuol dire *mucchio*). Metto un granello di sabbia sul tavolo. È un mucchio di sabbia? Ovviamente no. Due granelli? No. Dieci? Non ancora. Cento? Mille? Un milione? Se mille sì, perché non 999? Perché non a partire da 1001?

Variante: se una nuvola è un insieme di goccioline d'acqua più o meno vicine tra loro, dov'è esattamente il confine della nuvola, dove finisce la nuvola e inizia il cielo? Qualsiasi punto io indichi come confine mi chiedere potrà chiedere: e perché non un millimetro più a destra per includere nella nuvola anche quelle goccioline? Perché non un millimetro più a sinistra per escludere queste altre?

Questi esempi illustrano uno stesso conflitto. Da un lato, io ho il concetto di *mucchio*, che funziona in modo binario: cioè un certo ente o è un mucchio o non lo è. Dall'altro, l'ente che io indico con questo concetto è vago perché *complesso*, perché creato a partire da parti semplici: io individuo i due estremi, individuo l'insieme e la singola parte, il mucchio e il granello, ma tra i due ci sono molti gradi intermedi, e nessun granello ha in sé il potere di rendere subito mucchio un non-mucchio (questo lo posso dire anche per la nave con le sue assi, la nuvola con le sue goccioline, una sonata con le sue note, il mio corpo coi suoi atomi). Sorge così un'induzione indesiderabile:

1. aggiungo un granello ad un non-mucchio;
2. l'oggetto risultante è un non-mucchio;
3. quindi, se aggiungo un granello ad un non-mucchio, non posso mai ottenere un mucchio.

Questo è ovviamente assurdo, e lo è ancora di più quando lo metto al contrario: se tolgo un'asse da una nave, l'oggetto risultante è sempre una nave (non una non-nave), quindi, per quante assi possa togliere da una nave ne risulterà sempre una nave. E questo può essere detto per qualsiasi cosa abbia delle parti.

Il settore filosofico che si occupa della formalizzazione logica dei concetti di *tutto* e *parte* è la *mereologia*, e per me è una sorta di filosofia prima. Infatti questo potrebbe sembrare un problema piccolo piccolo, un esercizio intellettuale; e invece è il problema più importante di tutta la filosofia. Nelle sorite emerge tutto il conflitto tra il senso comune e la visione scientifica del mondo, tra il ragionamento *analogico* con infiniti livelli di specificazione e il ragionamento *digitale* che procede per categorie rigide, tra l'oggettivo e il soggettivo, tra l'apparenza e l'essere. Il problema delle sorite parla di come vengono individuati e distinti gli oggetti; parla del perché l'esistenza di una nave pare evidente, ma quando la si analizza scompare in un mucchio di assi distinte.

Un ente o è un mucchio o non lo è: come faccio a dire quando si dà un caso e quando l'altro? Se voglio mantenere questa binarietà del concetto, devo affermare l'esistenza di una soglia, un certo specifico numero di granelli che costituisce il mucchio. Secondo la teoria *costruttivista* tale soglia è creata come convenzione sociale, mentre secondo la teoria *epistemicista* la soglia è reale, dipendente dalla struttura del mondo e non dalle concettualizzazioni umane, ma sconosciuta. La maggior parte dei filosofi considera la prima soluzione come insoddisfacente, e la seconda come incredibile. Io, mettendo un paio di granelli su un tavolo, posso dire senza contraddizione che «la società crede che questo sia un mucchio, ma in realtà non lo è ancora», mentre sarei assurdo se dicessi «nella società occidentale non è ancora un mucchio, in quella orientale sì»: questo perché gli oggetti, per me, non sono costruzioni sociali, ma corrispondono ad una struttura reale del mondo indipendente dalle opinioni delle persone. Epperò, dall'altro lato, sembra impossibile individuare un fatto che possa costituire la soglia tra un mucchio e un non-mucchio di sabbia, nessun fatto naturale noto può svolgere un tale ruolo.

La terza possibilità sta nel rinunciare alla binarietà per passare alla *logica fuzzy*: non è vero che un oggetto o è un mucchio o non lo è, bensì ogni oggetto è *più o meno* un mucchio. Se metto un granello sul tavolo, quello è un mucchio all'1%; se continuo ad aggiungerne lo sarà al 10%, 30%, 70%, e via così. Questo vuol dire che la proposizione espressa da «questo è un mucchio» ha un valore di verità che aumenta man mano che aggiungo granelli, parte da zero (nessun granello, proposizione falsa) e può arrivare fino a uno (totalmente vera). Sono però ancora molti i filosofi che considerano la logica classica come sacra: dovrebbe essere l'ultima cosa che un filosofo va a toccare dopo aver provato con tutte le altre soluzioni.

Posso provare, allora, con una soglia mobile. Per esempio: se ho meno di mille granelli di sabbia sul tavolo, quando raggiungo i mille ho un mucchio; una volta raggiunta tale soglia, però, togliere un granello è insufficiente al tornare ad un non-mucchio, bensì devo andare sotto i 500. In altre parole, la frase «questo è un mucchio» è o vera o falsa, ma quale sia il suo valore di verità dipende dalla storia del mondo: 999 granelli non sono un mucchio se ho aggiunto un granello a 998, ma lo sono se ne ho tolto uno da mille.

Il problema, però, rimane sempre lo stesso: anche se la soglia è mobile, perché dovrebbe occupare una certa specifica posizione, e non una certa altra? Qual è il fatto che determina il posizionamento originario della soglia? Quine, con il suo approccio deflazionista, risponderebbe: non c'è alcun fatto del genere. Come il riferimento di «gavagai» è inscrutabile, così è il riferimento di «mucchio»: dicendo «mucchio» posso intendere 500 granelli, oppure posso (e dovrei) intendere *mucchio** cioè 499 granelli, o *mucchio*** cioè 501 granelli, la soglia può avere infiniti posizionamenti legittimi.

Questo può sembrare relativismo, e in effetti Geach propone (e poi critica) una teoria dell'*identità relativa*: se io ho una nave A composta da assi nuove, e una nave B composta dalle vecchie assi della nave di Teseo, allora la nave A è identica alla nave di Teseo nel senso che è la stessa nave (ma un diverso insieme di assi), e la nave B è identica alla nave di Teseo nel senso che è lo stesso insieme di assi (ma una diversa nave). Anche se tale interpretazione dell'identità è rifiutata dalla maggior parte dei filosofi, ha comunque il merito di evidenziare la differenza tra l'identità e la *composizione*.

Se perdo una mia parte, io divento la totalità meno quella parte: la composizione

La nave di Teseo è *composta* da assi, ma non è identica alle assi: per esempio, le assi hanno la proprietà di essere molte, la nave ha la proprietà di essere una, quindi tra le due cose non può esservi identità stretta. Di conseguenza la domanda sull'identità della nave non può essere solo una domanda sulla sua composizione: chiedermi se, nonostante il cambiamento, quella sia la stessa nave, non è un chiedermi se sia composta dalle stesse assi (lo stesso valga per mucchi e granelli, me e il mio corpo, eccetera).

Nella filosofia analitica c'è un esempio tradizionale per stimolare la riflessione sulla composizione, quello della statua e del marmo. Intuitivamente, il *David di Michelangelo* è identico in modo stretto al *marmo del David*: poiché sia il David, sia il suo marmo, sono solo un insieme di atomi in una certa struttura, e poiché i due concetti si riferiscono esattamente agli stessi atomi nella stessa struttura, non c'è alcuna differenza intelligibile tra i due. Eppure, se io prendo a martellate il David e lo riduco in polvere, il David cessa di esistere, ma il suo marmo no. Com'è possibile? Posso negare l'identità: posso dire cioè che il David e il suo marmo sono due enti diversi che occupano esattamente lo stesso spazio. Ma questo sarebbe assurdo. Allora devono essere lo stesso ente: perché uno viene distrutto dalle martellate, e l'altro no?

Anche se ora i riferimenti di *il David di Michelangelo* e di *il marmo del David* sono strettamente identici, questa è una contingenza, non una necessità: infatti i due concetti rimangono distinguibili in quanto si riferiscono a diversi insiemi di proprietà dello stesso oggetto (un po' come se parlassi del *colore della mela* e della *forma della mela* separatamente). Così, l'attuale David ha sia tutte le proprietà indicate da *il David di Michelangelo*, sia tutte le proprietà indicate da *il marmo del David* (è entrambe le cose); se io lo prendo a martellate fino a ridurlo in polvere, l'ammasso che ottengo avrà ancora tutte le proprietà de *il marmo del David*, ma nessuna delle proprietà de *il David di Michelangelo* (è ancora marmo, ma non più una statua); se, a suon di restauri, sostituisco uno a uno tutti gli atomi del David con atomi di alluminio, otterrò un oggetto che ha tutte le proprietà del *David di Michelangelo* ma nessuna delle proprietà del *marmo del David* (è ancora la stessa statua, ma non più un pezzo di marmo).

Si tratta di un'applicazione degli schemi mentali: c'è un unico oggetto, il quale è interpretabile sia come David, sia come marmo. Questo è ciò che succede anche con la nave di Teseo: un unico oggetto, interpretabile sia come nave (quindi un unico ente), sia come insieme di assi (quindi un insieme di enti). Essendo io innanzi a due concettualizzazioni di una stessa realtà, il confronto diretto è problematico: quando vedo la nave, non vedo un insieme di assi (se non come astrazione); e quando vedo un insieme di assi, non vedo la nave (se non come astrazione); poiché, se per me le assi sono oggetti singoli, non può esserlo anche la nave che compongono, e viceversa. La risposta alla domanda «se cambio questa asse, è ancora la stessa nave?» dipende dalla relazione esistente tra le proprietà indicate da *nave* e le proprietà indicata da *assi*: quando cambio un'asse, cambio delle proprietà indicate da *assi*, ma cambio anche delle proprietà indicate da *nave*? Similmente con il mucchio: l'aggiunta di un certo granello è sufficiente a creare la proprietà che voglio indicare con *mucchio*?

Quest'altro esempio, riproposto in tempi moderni da Geach, risale a Crisippo, filosofo stoico letteralmente morto dal ridere. Io ho un cane, si chiama Mozart. Quando dico «Mozart» intendo tutto lui, tutto ciò che è. Ora decido di chiamare «Moza» tutto ciò che lui è tranne la coda. In seguito ad un incidente con una capra, Mozart perde la coda. Chi c'è ora lì? Mozart è scomparso e c'è solo Moza? E prima, in quel corpo, c'erano sia Mozart e Moza, due entità che condividono lo stesso spazio?

Come il David col suo marmo, così *Mozart* e *Moza* si riferiscono a due diversi insiemi di proprietà, e la perdita della coda distrugge delle proprietà indicate da *Mozart*, ma nessuna delle proprietà di *Moza*. Come prendendo la statua a martellate distruggo il David ma non il suo marmo, così con la perdita della coda Mozart sparisce, ma non Moza.

Se questo lo posso dire a cuor leggero di una statua, con Mozart suona strano: è ovvio che un cane può ben sopravvivere al taglio della coda! Mozart è un insieme di cellule come la nave è un insieme di assi, ma non è *solo* un insieme di cellule: io credo che abbia l'esperienza di *un* soggetto, che abbia *un* suo punto di vista sul mondo, credo sia un individuo e non una pluralità di individui, e quindi neanche un ammasso di cellule una indipendente dall'altra. *Moza* indica una parte di Mozart, ma è solo un'astrazione, non è un individuo, Mozart esiste e Moza no. Così, quando Mozart perde la coda, è Moza che cessa di esistere in quanto diventa Mozart. Si rovescia la prospettiva: mentre nel caso di un mucchio di sabbia ho l'intuizione che il tutto (il mucchio) sia un'astrazione e solo le parti (i granelli) esistano davvero, con un essere vivente ho l'impressione che esista il tutto ma non le parti.

C'è infatti una differenza notevole tra *Mozart* e *Moza con la coda*: *Moza* è una parte di Mozart, e così la sua coda, e Mozart è contingentemente identico a *Moza con la coda* proprio come il David lo è al suo marmo e la nave di Teseo all'insieme delle sue assi; ma *Moza con la coda* concettualizza un insieme contenente due enti, può indicare proprietà assegnate a due enti diversi, mentre Mozart, come la nave, concettualizza un unico ente. Posso allora scrivere così:

- le parti di Mozart, se non relazionate, hanno esistenza autonoma;
- quando queste parti entrano tra loro in una certa relazione, si fondono in un unico ente i cui sottoinsiemi sono astrazioni che non costituiscono ulteriori entità (se il concetto *Mozart* indica una qualche proprietà, questa potrebbe emergere dalla relazione delle parti);
- le parti possono riacquisire esistenza autonoma distaccandosi da questo ente, e altre possono perderla fondendosi in esso, ma l'ente continua ad esistere (almeno finché le proprietà alle quali il concetto *Mozart* si riferisce persistono).

Se limito tale teoria agli esseri coscienti come Mozart, allora ottengo un *nichilismo parziale* (o *ristretto*, o *organicista*) come quello del filosofo Trenton Merricks: gli esseri viventi esistono (pur composti da cellule), mentre oggetti inanimati come la nave di Teseo sono astrazioni in quanto solo le loro parti esistono veramente.

Volendo invece portare avanti il paragone ed estendere la teoria dagli esseri viventi alla nave di Teseo, dirò che: la nave, come Mozart, è un unico ente, e le assi, come Moza, sono sue parti (ma non enti reali), quindi, man mano che sostituisco le assi, la nave sarà contingentemente identica ad un diverso insieme di parti (almeno finché si mantiene una certa relazione tra le assi). Al che viene naturale sostenere (anche se non è necessario) che nel problema delle sorite, una volta che i granelli formano il mucchio, questo deve restare tale anche sottraendo alcuni granelli, proprio come vuole la teoria della soglia mobile.

Né le proprietà corporali né quelle psicologiche sono essenziali

Ma allora, quali sono le proprietà alle quali tento di riferirmi quando dico «io»? Cioè, poiché io ho continuità nel tempo, cos'è che mi rende sempre me stesso nonostante i tanti cambiamenti nelle cellule che mi compongono? La filosofia analitica ha proposto due risposte principali: per la prima,

la continuità della persona è garantita dalla *continuità corporale*, per la seconda, dalla *continuità psicologica*.

La teoria della continuità corporale non può essere generalizzata a tutto il corpo: io potrei perdere braccia e gambe, potrei aver bisogno di un trapianto di cuore, polmoni, reni e fegato, eppure sarei sempre io. È vero però che non posso subire un trapianto di cervello, e che i miei neuroni sono cambiati meno rispetto alle cellule di altre parti del mio corpo, la mia pelle cambia di continuo ma i miei neuroni no. Quindi potrei sostenere che, se mi tagliassero la testa, o qualcuno estraesse il cervello dal mio cranio, io sarei lì dov'è il mio cervello, e non dov'è il resto del corpo. Può il cervello garantire la mia continuità nel tempo?

Due particelle elementari possono essere qualitativamente identiche, posso immaginare due elettroni identici in tutto tranne che per il loro posizionamento nello spazio. Posso prendere un atomo, e immaginare di sostituire un suo elettrone con un altro qualitativamente identico: nulla di rilevante cambierebbe in quell'atomo, sarebbe sempre lo stesso. Quindi continuo fino a sostituire tutte le particelle di tutti gli atomi di tutti i miei neuroni. È il *cervello di Teseo*: alla fine ottengo un organo strutturato in modo identico a quello di partenza, ma composto da materia diversa. Sarei sempre io? La risposta intuitiva è sì: posso ben immaginare, infatti, che gli atomi del mio cervello guadagnino o perdano qualche elettrone di continuo, così come che i miei neuroni guadagnino o perdano qualche atomo nel corso della giornata, e ciò di sicuro non distrugge il mio *io*. In altre parole: la materia è neutrale, ma la sua struttura no, anche se cambio la materia il fatto che il cervello abbia la stessa struttura implica la mia identità.

Peccato che tale risposta sia sbagliata: infatti io so che la struttura del mio cervello cambia di continuo, ci sono neuroni che muoiono, altri (pochi) che nascono, sinapsi che si rafforzano, altre che si indeboliscono, eppure sono sempre io. Così, da un lato, la materia, essendo neutrale, non può costituire un fatto giustificante la mia continuità nel tempo; dall'altro, è evidente che neanche la struttura può farlo in quanto cambia di continuo senza che l'io cambi. E visto che il mio corpo è null'altro che materia e struttura, l'ipotesi della continuità corporale risulta difficile da difendere.

Infatti molta più fortuna ha avuto l'ipotesi della continuità psicologica, analizzata da alcuni dei più grandi filosofi del secolo scorso come Parfit, David Lewis, e Nagel. La variante più nota di tale teoria è quella che riduce la continuità personale alla continuità della memoria: se io mi sveglio una mattina ricordando di aver passato la sera prima ad una festa, dirò che sono io quella persona che ieri ha festeggiato; ma se le mie memorie si fermassero ad una settimana prima, e un amico mi dicesse di avermi visto alla festa, io risponderei che, chiunque fosse quella persona, non ero io, era un corpo inconscio o controllato da un'altra mente.

Posso però immaginare subito un controesempio. Immagino che io non sia andato alla festa, ma il mio vicino di casa sì, e che mentre dormo un demone manipoli il mio cervello in modo tale che io acquisisca tutti i ricordi del mio vicino di casa. Questo mi renderebbe forse lui? Sarei io quello che è andato alla festa? Certo che no, sarebbe assurdo immaginare il vicino che prende possesso del mio corpo solo in virtù della condivisione dei ricordi. Posso ripetere l'operazione con gli interessi, il comportamento e ogni altra caratteristica psicologica, ma nulla di ciò è sufficiente a rendermi un'altra persona. Posso immaginare anche di sviluppare devastanti malattie mentali, posso ammalarmi di Alzheimer e avere continue amnesie, tutta la mia psiche cambierebbe in modo radicale, eppure sarei sempre io, e se qualcuno dicesse «non è più lui!» sarebbe solo retorica.

Derek Parfit fa un esempio che prevede sia continuità corporale che psicologica. Ho un teletrasportatore in stile Star Trek: entro nella cabina di partenza, la posizione di tutte le mie

particelle viene memorizzata, il mio corpo viene distrutto mentre, nella cabina di destinazione, la mia struttura corporea viene ricreata atomo per atomo. Poiché in ogni istante c'è un solo individuo con le mie caratteristiche fisiche e psicologiche, l'intuizione è che io sia scomparso da una parte e ricomparso dall'altra. Un giorno, però, il teletrasportatore ha un malfunzionamento: ha copiato la posizione dei miei atomi e l'ha ricreata nella cabina di destinazione, ma non ha distrutto il mio corpo, l'ha solo destabilizzato e si smolecolarizzerà da solo tra qualche ora. Ora che i corpi con le stesse caratteristiche sono due devo chiedermi: dove sono io? Perché dovrei dire di essere nella cabina di destinazione, se non sono mai sparito dalla cabina di partenza? È così: il corpo apparso nella cabina di destinazione è identico al mio sia in quanto a struttura materiale che a psicologia, ma non sono io; quindi né il corpo, né la psiche sono sufficienti a giustificare l'identità di una persona con sé stessa.

Un altro esperimento fantascientifico si trova in Bernard Williams. Ho una macchina in grado di scambiare la psiche di due persone, le addormenta e ristrutturata i loro cervelli in modo tale che quello di uno diventi identico a quello dell'altro. Ci sono due volontari, Abelardo ed Eloisa. Spiego ad Abelardo che scambierò la sua psiche con quella di Eloisa, poi darò una torta ad uno dei due corpi, ma deve essere lui a decidere quale. Cosa sceglierà, se è furbo? Dopo l'esperimento, il corpo di Abelardo avrà la psiche di Eloisa, e il corpo di Eloisa la psiche di Abelardo. Abelardo, che vuole mangiare, deve scegliere di darla al corpo di Eloisa in quanto quello sarà il corpo che desidererà per sé la torta come lui la desidera adesso. Cioè Abelardo sposa la teoria della continuità psicologica.

Immagino però un altro caso. Dico ad Abelardo che la sua psiche sarà scambiata con quella di Eloisa, e poi il suo corpo, quello che avrà la psiche di Eloisa, verrà picchiato a sangue. Dovrebbe essere preoccupato di ciò? Pare proprio di sì: infatti, come il desiderio di mangiare continuerà ad essere diretto verso la torta qualsiasi sia il corpo che supporta quella psiche, così il desiderio di integrità fisica di Abelardo sarà diretto verso il suo corpo anche dopo lo scambio; quindi, non volendo essere picchiato, chiederà di fare picchiare il corpo di Eloisa, nonostante la psiche di quel corpo sia come la sua. In questo caso, allora, Abelardo passa alla teoria della continuità fisica.



Figura 11

Per individuare quale sia l'intuizione più forte, ora chiedo ad Abelardo di scegliere sia quale corpo deve ricevere la torta, sia quale deve essere picchiato. Posto che non può scegliere Eloisa in entrambi i casi, è probabile che Abelardo sia preoccupato dell'evitare il pestaggio più dell'aver la

torta: per questo Williams crede che l'intuizione della continuità corporale sia più forte di quella della continuità psicologica. Resta però il fatto che nessuno dei due criteri è sufficiente a giustificare la continuità personale: se uno è preferibile in certi casi, l'altro lo è in altri.

Quindi né la teoria della continuità corporale, né quella della continuità psicologica, sono soddisfacenti; faccio allora un'ultima prova con una teoria mista. In primo luogo, riconosco la dipendenza della continuità psicologica dalla continuità corporale: è infatti normalmente accettato che i propri stati psicologici dipendano dagli stati del cervello, cioè che non vi sia continuità psicologica senza una certa continuità corporale. In secondo luogo, riconosco che lo stato attuale del mio cervello è causato dai suoi stati passati, cioè il mio cervello evolve nel tempo in accordo con il controfattuale «se non si fosse dato quello stato passato, non si sarebbe dato questo stato presente». Di conseguenza: ogni volta che, indicando due enti, dico «sono io» con verità, esiste una connessione causale che da uno porta all'altro.

Quindi: io sono lo stesso individuo che ero da bambino se io sono come sono *perché* quel bambino era com'era; l'io che si è svegliato stamattina e l'io che si è addormentato ieri sera sono lo stesso individuo se io mi sono svegliato in un certo stato *perché* mi sono addormentato in un certo altro. Se l'amico mi fa una foto mentre dormo e poi me la mostra, posso ben dire «sono io» anche se in quel momento non avevo stati psichici, quindi la sola continuità psicologica non può essere sufficiente, l'*io* non è la coscienza; ma la continuità psicologica si basa su una continuità corporale che porta dalla veglia al sonno e viceversa, e in virtù di ciò sono sempre io anche mentre dormo.

Questa teoria, per essere accettabile, deve rispondere ad una domanda molto difficile. Io non sono eterno, il mio *io* può cessare di esistere, il che accade, secondo tale teoria, quando la catena causale tra i miei stati si interrompe. Ma: cosa causa la rottura di questa catena?

Di sicuro il mio cervello può cambiare a causa di interventi esterni senza che ciò distrugga l'*io*. Il mio cervello è costantemente bombardato da raggi cosmici che ne modificano la struttura atomica, può subire piccoli danni, può cambiare in seguito a interventi chirurgici, posso addirittura finire in stato vegetativo e perdere la capacità di generare nuovi stati psicologici coscienti, ma io rimango sempre io, quindi continua ad esserci un rapporto causale tra i miei stati. Parfit immagina allora il caso più traumatico possibile, quello dell'*emisferectomia*, un intervento chirurgico che può prevedere l'estrazione di un intero lobo cerebrale, metà cervello. Per incredibile che sembri, i pazienti sopravvivono e riescono a recuperare pressoché tutte le funzionalità perse, a volte anche ad un livello superiore alla media. Questa è la realtà, ora la fantasia: a causa di un istante di confusione del comitato etico dell'ospedale, viene permesso ad un chirurgo di asportare il mio emisfero cerebrale destro per metterlo nel cranio vuoto di un altro corpo. Dopo l'intervento, dove sarò io?

La prima possibilità è che entrambi gli emisferi mantengano il rapporto causale tra i loro stati: continua infatti ad essere vero che il loro stato presente sarebbe stato diverso con diversi stati passati. Questo però implica che l'individuo che possiede un certo emisfero dopo l'intervento sia lo stesso che lo possedeva prima: cioè, dopo l'intervento io sono in entrambi i corpi, ci sono due corpi ma un'unica persona. Questo è assurdo.

La seconda possibilità, allora, è che l'intervento sia troppo radicale, e abbia distrutto il rapporto causale per entrambi gli emisferi: quindi io ho cessato d'esistere, e i due corpi costituiscono due nuovi individui. Ma l'*emisferectomia* è un intervento fatto per salvare la vita di un individuo, non per crearne di nuovi: se l'emisfero destro fosse stato estratto e distrutto invece che trapiantato, chi sarebbe sopravvissuto all'intervento? Io, ovviamente. Allora anche questa interpretazione suona inaccettabile.

L'ultima possibilità è che io continui ad esistere nell'emisfero sinistro che non ha mai lasciato il cranio, mentre il destro, venendo spostato in un altro corpo, genera una nuova persona. Di contro, però, è intuitivo che, se avessi trapiantato l'intero cervello, l'individuo risultante, pur cambiando corpo, sarei stato sempre io. È quindi vero che il trapianto in un altro corpo non è sufficiente a rompere il collegamento causale tra gli stati: perché allora l'emisfero sinistro dovrebbe mantenere tale collegamento, ma quello destro no? Qualsiasi sia il motivo, non è tra quelli contemplati da questa teoria, che risulta quindi insoddisfacente.

Ma se né il corpo né la psiche possono costituire l'io, e in me ci sono solo queste due cose, cos'altro può farlo? Hume risponde: nulla. Il tutto sta alla parte come la nave alle assi, il corpo alle cellule, le esperienze mentali all'io. Come posso concettualizzare un certo ente sia come nave che come insieme di assi, così posso concettualizzare la mia psiche sia come *io* sia come *insieme di esperienze*. Ma poiché le esperienze esistono in modo evidente, mentre non esiste alcun fatto che costituisca la soglia in merito a quante esperienze possa cambiare prima di cambiare l'io, posso abduire che non esista neanche un fatto che costituisce l'io, sto tentando di riferirmi a proprietà che non ci sono: esistono insiemi di esperienze, ma non c'è alcun *io* che fa tali esperienze; corollario, io non ho continuità nel tempo, in ogni istante sono un diverso ammasso di esperienze, e se quando dormo non ho esperienze non sono neanche quello.

Questa teoria humeana, vagamente buddista, è però troppo lontana dal senso comune per essere accettata così: posso negare che esista un fatto che costituisce il riferimento delle parole, ma non posso concludere che non esiste il significato, devo trovare una soluzione scettica; similmente, posso negare che esista un fatto che costituisce la mia continuità nel tempo, ma non posso concludere che non vi sia alcun io che continua nel tempo, devo trovare una soluzione scettica.

Neanche la mia origine è essenziale

Ora dico: «invece che un filosofo, sarei potuto essere un architetto». Dico cioè che esistono due mondi possibili: nel primo faccio la vita che ho fatto, e finisco per essere filosofo; nel secondo, invece, prendo scelte alternative, sviluppo una mente diversa, un diverso corpo, e finisco per essere un architetto. La domanda è: l'*io-filosofo* e l'*io-architetto* sono la stessa persona? Poiché sto descrivendo due possibili percorsi della mia vita, è così per definizione, sono sempre io.

Questo tipo di *identità transmondana* (cioè tra mondi possibili) richiede qualcosa di diverso dalla connessione causale tra gli stati. Io tiro un dado, esiste un mondo possibile in cui esce tre e uno in cui esce sei, e il *dado-tre* e il *dado-sei* sono lo stesso dado; ciò nonostante, tra i due non c'è alcun collegamento causale, non è vero che uno stato causa l'altro. Io potrei accettare di essere identico al mio corpo *vivo*, potrei dire che la mia continuità nel tempo è giustificata dalla connessione causale esistente tra gli stati delle mie funzioni biologiche; ma tra il corpo dell'*io-filosofo* e il corpo dell'*io-architetto* non c'è connessione causale, sono due storie del mondo parallele, una indipendente dall'altra. Allora, cos'è che rende l'*io-filosofo* e l'*io-architetto* la stessa persona?

Il perdurantismo mi concepisce come una catena causale di stati: il mio io bambino ha causato il mio io adolescente che ha causato ciò che sono adesso. Visualizzo questo come un *serpente* che si estende nel tempo: io non sono questo corpo e basta, ma anche tutti i corpi che sono stato in passato. Anzi, ancora di più: io sono tutte le vite alternative che avrei potuto fare, tutti i serpenti possibili che avrei potuto percorrere, tutti i futuri tra i quali posso scegliere. Cioè, io sono fatto così, sono esteso nello spazio, nel tempo, e anche in una dimensione modale:

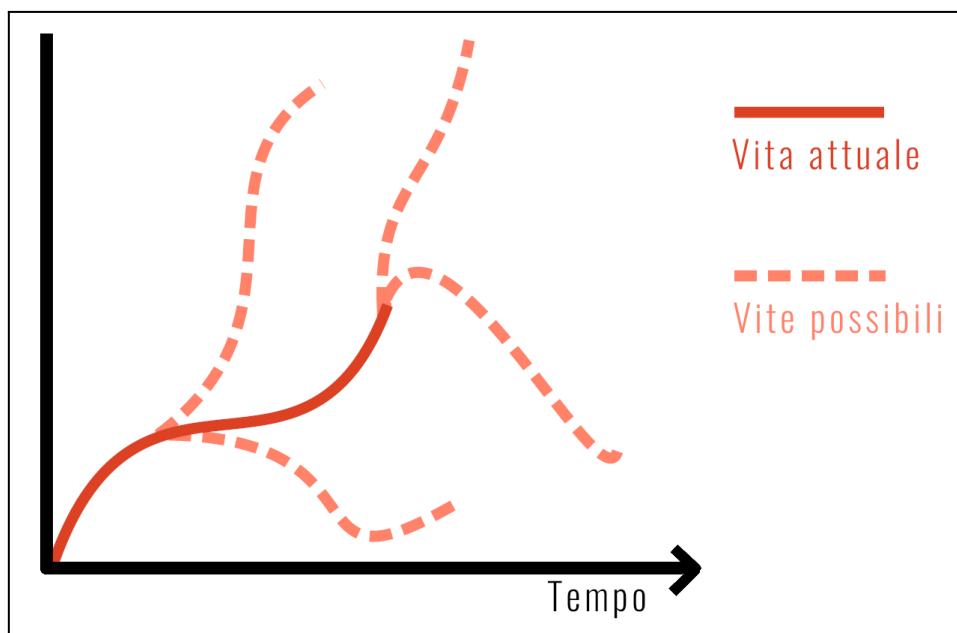


Figura 12

L'endurantismo, invece, mi chiede di cercare una *essenza*, una proprietà senza la quale io non sarei io e che deve dunque essere presente in tutti i miei stati senza mai cambiare. Questo è complementare alla visione perdurantista: deve infatti esservi comunque un fatto in virtù del quale l'io delle varie vite possibili è sempre lo stesso. Ma cosa può esservi di comune ad un infinito numero di vite diversissime? Cioè: qual è quel fatto che, qualsiasi vita io abbia deciso di condurre, non può non caratterizzarmi?

Rispondo: la mia *origine*. È vero, avrei potuto fare una vita diversa, avrei potuto prendere diverse decisioni, e questo sin dal primo istante della mia esistenza... *ma non prima!* Infatti, prima di esistere, non potevo certo compiere scelte. Così, tutte le mie vite possibili devono avere un punto in comune, che è il primo istante della mia esistenza. Ciò che rende tutti i miei *io* possibili lo stesso identico individuo è la possibilità di risalire, da qualsiasi loro stato, alla stessa origine; e perciò solo l'origine, il punto dal quale tutte le catene causali partono, può essere quella proprietà che, se modificata, mi rende una persona diversa.

La prima domanda che mi faccio è cosa sia esattamente questa mia origine. Kripke, con una mossa più immanente di quanto sia sua abitudine, sembra pensare ad un qualche evento biologico. Un certo spermatozoo ha fertilizzato un certo ovulo, è diventato uno zigote, ha iniziato a dividersi e a formare una morula. A questo livello io non esisto ancora perché la morula è solo un ammasso di cellule *totipotenti*, avrei potuto separarle per formare infiniti individui possibili; poi, però, le cellule hanno perso questa capacità, si è formata la gastrula e da allora quell'ente non può più essere un altro, sono diventato io e lo sono ancora, e lo sarò, indipendentemente da quante mie proprietà possono cambiare, finché mi sarà possibile costruire una storia che, da me, torna indietro fino a quella gastrula.

La seconda domanda riguarda invece la necessità dell'identità. È chiaro che io sono *necessariamente* identico a me stesso, cioè non esiste alcun mondo possibile nel quale io posso essere diverso da me stesso. In altri casi pare però possibile un'identità contingente: ho detto che il David è solo *contingentemente* identico al suo marmo. È così? Ruth Barcan Marcus, una delle grandi logiche del

secolo scorso, crede di no, per lei ogni relazione d'identità è necessaria. La sua argomentazione è questa:

- Tizio=Caio;
- Tizio è necessariamente identico a Tizio stesso;
- se qualcuno è necessariamente identico a Tizio, questo è Tizio (e nessun altro);
- ma poiché Tizio=Caio, posso scambiarli per ottenere: se qualcuno è necessariamente identico a Tizio, questo è Caio;
- quindi Tizio è necessariamente identico a Caio.

Gareth Evans, geniale filosofo critico di Kripke morto a 34 anni dopo aver dato inestimabili contributi alla teoria del riferimento, mostra inoltre che il valore di verità di un'identità non può essere indeterminato, deve essere o vera o falsa:

- Tizio=Caio;
- ipotizzo che il valore di verità di «Tizio=Caio» sia indeterminato;
- il valore di verità di «Tizio=Tizio» è sicuramente determinato;
- se qualcuno è determinatamente identico a Tizio, questo è Tizio (e nessun altro);
- ma poiché Tizio=Caio, posso scambiarli per ottenere: se qualcuno è determinatamente identico a Tizio, questo è Caio;
- quindi il valore di verità di «Tizio=Caio» è determinato (e se è indeterminato, è falso).

Pare logico: la statua e il marmo sono *contingentemente* identici perché non sono identici in primo luogo, sono concetti che indicano entità diverse. Di conseguenza, se voglio dimostrare che l'origine è essenziale e quindi la base dell'identità, cioè che essere identici vuol dire avere la stessa origine, allora devo dimostrare anche che l'origine è necessaria e determinata, cioè che è impossibile avere un'origine diversa da quella che si ha.

Kripke tenta di dimostrare che, in effetti, la necessità dell'identità implica l'essenzialità dell'origine. Nel mondo attuale ci sono due tronchi, che chiamo *A* e *B*. Da *A* ricavo un tavolo, che chiamo *C*. Per il *principio di composibilità* esiste un mondo possibile nel quale, oltre a *C* da *A*, ho ricavato anche il tavolo *D* dal tronco *B*. Per l'*unicità dell'origine* è impossibile che uno stesso tavolo abbia due origini diverse contemporaneamente, quindi in questo mondo possibile i tavoli *C* e *D* devono essere distinti, non numericamente identici. Per il *principio di sufficienza*, se in un mondo possibile io ricavo solo un tavolo, dal tronco *B*, questo sarà sempre *D* e cioè sempre diverso da *C* (come l'identità è necessaria, così è la disuguaglianza); quindi non esistono mondi possibili nei quali posso ricavare *C* da un tronco diverso da *A*.

Questa argomentazione non convince molti. Se l'essenzialità dell'origine è vera, lo è anche l'essenzialità della composizione: è vero che, *temporalmente*, tutto il marmo che compone il David può cambiare, ma, *atemporalmente*, il David non poteva originare da un altro blocco di marmo né da un diverso materiale. Michael Rea, epistemologo cristiano, non è d'accordo: essendo il David di Michelangelo diverso dal suo marmo, la stessa statua poteva benissimo essere creata da un diverso blocco. David Lewis, similmente, accetta a cuor leggero controfattuali che riguardano l'origine come «invece che nascere da uno zigote, poteva portarmi la cicogna» oppure «invece che in Italia nel 1993, sarei potuto nascere in Cina nel medioevo»: è legittimo dire che avrei potuto avere origini diverse, quindi l'origine non è necessaria, e quindi non può essere la base dell'identità.

Questo contrasto origina in un disaccordo sulla semantica. Prendo il controfattuale «sarei potuto essere dottore»: Kripke lo interpreta in modo letterale come «esiste un mondo possibile nel quale

io (*proprio io*) sono dottore», David Lewis invece lo interpreta come «esiste un mondo possibile nel quale una mia controparte (e non io *io*) è dottore». Di conseguenza, il controfattuale «io e il mio corpo saremmo potuti essere diversi» è assurdo in Kripke (pur accettando la distinzione tra me e il mio corpo), mentre Lewis lo accetta senza problemi in quanto lo interpreta come «esiste un mondo possibile in cui c'è una mia controparte personale e una mia controparte corporale tra loro diverse». Questo potrebbe essere esteso anche ai controfattuali in cui si predica, per un certo ente, una natura diversa dalla sua, come in «se mia nonna avesse le ruote sarebbe una carriola»: queste sono possibilità logiche, è logicamente possibile che mia nonna sia una carriola, che un oggetto sia rosso ma non esteso nello spazio, che l'acqua non sia H₂O; ma se siano anche possibilità metafisiche, alcuni filosofi lo credono, altri lo negano.

Quindi nessuna proprietà è in me essenziale: una prospettiva minimalista

L'essentialismo risulta quindi poco convincente. Il *massimalismo*, cioè l'idea che ogni proprietà sia essenziale, che nessuna proprietà possa essere modificata senza negare l'identità, lo è ancora meno. L'unica alternativa è il *minimalismo*, l'idea che nessuna proprietà sia essenziale, e non vi sia quindi limite a ciò che è modificabile.

Penso che il rapporto esistente tra il mio corpo, la mia psiche, e l'io, sia simile a quello esistente tra il David, il suo marmo, e l'ente che questi concetti indicano.

Se prendo a martellate il David, ottengo un ammasso di materia per il quale non posso più affermare «questo è il David», ma posso ancora dire «questo è il marmo del David»: ho quindi distrutto la proprietà essenziale del David, ma non la proprietà essenziale del marmo, che deve stare tra quelle al di fuori dell'insieme di proprietà indicato dal concetto *David*. Al contrario, se sostituissi la materia del David atomo per atomo, distruggerei l'essenza del marmo, ma non quella del David.

Ora ripeto questo ragionamento applicandolo a me stesso. Provo a dire: «io avrei potuto avere un corpo diverso». Se è così, tutte le proprietà del mio corpo possono cambiare o essere cancellate senza che ciò mi renda una persona diversa. In virtù degli esempi visti, questo sembra vero, e concludo allora che la proprietà essenziale che mi rende *me* deve stare al di fuori dell'insieme delle proprietà del corpo. Sta forse tra le proprietà della psiche? Posso affermare anche «io avrei potuto avere una psiche diversa», cioè posso cambiare o cancellare anche tutte le proprietà della psiche, tutte le mie esperienze coscienti e i miei pensieri, senza cambiare io: e anche questo sembra vero in virtù degli esempi visti.

Quali altre proprietà posso avere? Posso cancellare *contemporaneamente* le proprietà del corpo e della psiche, senza cancellare la proprietà indicata dall'*io*? È fortemente controintuitivo. Come l'ente che indico con *il David di Michelangelo* e *il marmo del David* ha entrambi gli insiemi di proprietà ma null'altra in più, così l'insieme *io* è l'unione degli insiemi di proprietà de *il mio corpo* e *la mia psiche*, e non ho ulteriori proprietà. Ma se è così, pur essendovi proprietà *triviali* che per logica devo avere (come l'essere identico a me stesso), comunque nessuna proprietà *non-triviale* mi è essenziale. Il corpo non mi è essenziale, la psiche non mi è essenziale, ma se elimino corpo e psiche non rimane neppure l'io. Se il mio corpo cambia, io divento quel nuovo corpo, viene inglobato dall'*io* in virtù della mia continuità psichica; e se poi anche la mia psiche cambia, io divento quella nuova psiche, viene inglobata dall'*io* in virtù della continuità del nuovo corpo; cioè, ogni singola proprietà può cambiare, basta che non ogni singola proprietà cambi nel medesimo istante.

Questa posizione è in linea con quella di Hume. È anche la posizione di Parfit, che fa l'esempio della Nazione: una Nazione è irriducibile al suo territorio, è irriducibile al suo popolo, eppure è null'altro sopra queste due cose.

Infine, è anche una posizione vicina alla teologia cristiana: per San Tommaso, ad esempio, la persona non è il suo corpo, non è neppure la sua anima, eppure è null'altro che corpo e anima (*unità*). L'idea che l'identità non dipenda da alcuna essenza porta al problema della *pura identità*, il fatto che in un altro mondo possibile Elvis e Nixon possono avere ognuno tutte le proprietà dell'altro, di modo che l'unica cosa che cambi in quel mondo sia l'identità dei due insieme: invece della star Elvis invitata dal presidente Nixon vi sarà la star Nixon (chiamata Elvis) invitata dal presidente Elvis (chiamato Nixon). Anche se questa visione causa disagio in molti filosofi, alla fin fine è solo una riedizione moderna dell'*eccettismo* del francescano Duns Scotto, e anche se David Lewis lo rifiuta credo possa essere esposto in un modo compatibile con la sua semantica.

Che l'identità in generale non possa essere sempre spiegata tramite la presenza di proprietà essenziali è intuitivo. Infatti, io posso produrre affermazioni anche sull'identità delle proprietà stesse: posso dire che l'essere buoni è identico al fare il proprio dovere, per esempio. Ma in base a quale fatto le due proprietà risultano identiche? Sarei assurdo se dicessi che le proprietà stesse hanno proprietà essenziali. Devono allora avere una *eccità* non-qualitativa che giustifica l'identità delle proprietà medesime: non è un'essenza, è solo una base dell'identità.

Faccio ora un esempio con l'emisferectomia per evidenziare le conseguenze di questa idea sull'identità personale:

- ci sono due corpi, *A* e *B*; il corpo *A* sono io; il corpo *B* è identico al corpo *A* con la sola differenza che non ha un cervello;
- se sposto l'emisfero destro del corpo *A* nel corpo *B*, creo un individuo chiamato Mario;
- se sposto l'emisfero sinistro del corpo *A* nel corpo *B*, creo un individuo chiamato Luigi;

se io ora sposto l'emisfero destro nel corpo *B*, chi è il corpo *A*?

Il corpo *B* è Mario. Intuitivamente, il corpo *A* devo essere io: sono cambiate alcune sue proprietà non essenziali (ha perso un emisfero), quindi la sua identità personale deve essere rimasta la stessa. Ma: ora cosa distingue questo corpo *A* da Luigi? Niente, è lo stesso emisfero cerebrale in un identico corpo. È vero che Luigi (cioè l'emisfero destro in *B*) ha una storia diversa rispetto a me (che sono l'emisfero destro in *A*): Luigi è prodotto da un trapianto, da uno spostamento dell'emisfero, mentre nel mio caso questo non è avvenuto. Ciò nonostante, questa proprietà non è essenziale, Luigi potrebbe avere una storia identica alla mia e rimarrebbe sempre sé stesso (e questo valga per qualsiasi altra differenza si possa evidenziare).

Questo significa, in primo luogo, che l'identità personale è un concetto distinto dall'identità numerica rappresentata dalla legge di Leibniz; e, in secondo luogo, che le persone non sono insieme di proprietà perfettamente distinti uno dall'altro, sono bensì sovrapposte e in alcuni casi addirittura coincidenti. È come se le varie vite possibili siano dei ruoli da interpretare, e questi possano essere assegnati a diversi attori, ma una volta che io ottengo una parte (cioè ho una certa vita attuale), nessun altro attore può averla, anche se l'altro attore potrebbe avere la parte che io avrei avuto agendo diversamente. In altre parole: io sono l'insieme dei miei stati possibili, e c'è una scia di stati attuali che sono solo miei (l'io-filosofo, dico), ma i miei stati possibili possono essere gli attuali di qualcun altro, il mio io-architetto possibile potrebbe essere l'attualità di qualcun altro, e quindi io e questo architetto avremmo una *parte modale* in comune (ma mai parti attuali).

Così, quando indico il corpo A chiedendomi «chi è quello?» la mia risposta varia a seconda di quali sono le *altre* persone che ritengo essere presenti nel mondo. Questa posizione è totalmente epistemologica, non ontologica: non esiste davvero una dimensione modale, ed esiste invece un fatto in merito all'identità ma, nell'impossibilità di conoscerlo, devo affidarmi alle mie intuizioni. In questo mondo possibile esiste Luigi, oppure no? Se esiste, allora il corpo A è lui, e io non esisto; se non esiste, allora il corpo A sono io. Dal punto di vista epistemico, se vi sia identità tra me e il corpo A , e in virtù di quali proprietà, dipende non solo dal modo in cui concettualizzo me stesso (se mi considero appunto un corpo, o invece un'anima, o un ente quadridimensionale...), ma dalla visione del mondo in generale.

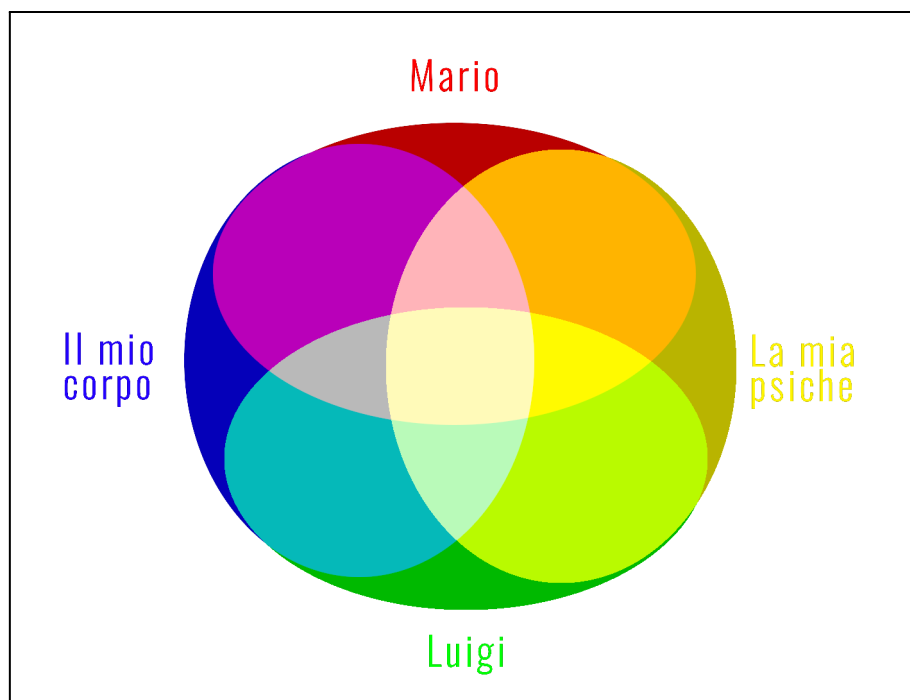


Figura 13

L'eternalismo non rende merito di tutti i fatti: galassie aliene e proprietà intrinseche

Tutte le soluzioni proposte per il problema delle sorite e della nave di Teseo sono vicine all'endurantismo, e quindi al presentismo. Le soluzioni per l'identità transmondana, come l'essenzialismo dell'origine (che è anche un'altra soluzione per la nave di Teseo), funzionano invece meglio presupponendo il perdurantismo, e quindi l'eternalismo. Analizzare le teorie del tempo può quindi aiutare a comprendere quale teoria dell'identità risulti essere più convincente. E allora inizio analizzando le ragioni della teoria eternalista.

Putnam, che aveva le mani in pasta in ogni campo della filosofia, in squadra con i fisici Rietdijk e Penrose (il quale ha vinto il Nobel durante la revisione di questo capitolo), ha sviluppato un'importante argomentazione a difesa dell'eternalismo. Ci sono più versioni di questa *argomentazione di Rietdijk-Putnam-Penrose*, quella di Penrose nello specifico si chiama *argomentazione Andromeda*, ed è quella che espongo di seguito.

Il Galaxy Express 999, un treno che compie viaggi intergalattici, è da poco partito dalla Terra e si sta dirigendo ad altissima velocità verso la galassia Andromeda. Strada facendo incrocia la nave di Capitan Harlock, che sta viaggiando ad una diversa velocità da Andromeda verso la Terra. La nave e il treno, una di fianco all'altro, condividono un certo istante presente, che chiamo t . Dalla teoria della relatività so però che le due astronavi, viaggiando a velocità diverse, hanno anche diversi *piani degli eventi presenti*: quali siano gli altri eventi dell'universo simultanei al loro incontro varia a seconda che si stia viaggiando nella nave o nel treno. Infatti, poiché al variare della velocità varia anche il modo in cui scorre il tempo, nell'istante t , mentre dalla prospettiva della nave gli alieni di Andromeda stanno discutendo se invadere la Terra o meno, dalla prospettiva del treno gli alieni sono già partiti e pronti a invadere questo pianeta. Il *piano degli eventi presenti* è interpretabile come una sezione tridimensionale dell'universo, cioè un suo istante atemporale, ma poiché in questo stesso universo esistono più *piani degli eventi presenti* (ogni ente che si muove ha il suo), è impossibile che l'universo nel suo complesso sia tridimensionale, è impossibile che sia esso stesso *un* piano; deve dunque essere almeno quadridimensionale, esteso nel tempo, quindi l'eternalismo deve essere vero.

Molti fisici hanno criticato questa argomentazione: si ritiene infatti che il concetto di *presente* nella teoria della relatività sia locale e non generalizzabile a tutto l'universo, cioè posso parlare di eventi simultanei se questi sono vicini (come l'incrocio della nave col treno), ma non posso mettere in relazione il presente di queste astronavi con degli eventi nella lontana Andromeda, non posso dire legittimamente quali eventi di Andromeda sono simultanei ad un evento che riguarda il treno.

David Lewis difende l'eternalismo con una diversa, potente strategia. È normalmente accettato che esistano *proprietà intrinseche*. Le proprietà intrinseche sono quelle non-relative, cioè le proprietà che un oggetto avrebbe anche se non esistesse altro nell'universo: per esempio, il peso è una proprietà relativa perché dipende dalla gravità, la massa invece è intrinseca; infatti, posto un oggetto, se voglio cambiare il suo peso posso limitarmi a modificare le sue relazioni con l'ambiente (posso portarlo sulla luna, per esempio), ma se voglio cambiare la sua massa sono costretto a modificarlo in modo diretto. Ora: io oggi ho i capelli lunghi, domani li taglierò e li avrò corti. Se *avere i capelli lunghi* e *avere i capelli corti* fossero proprietà a me intrinseche, sarei un essere contraddittorio (avrei i capelli contemporaneamente lunghi e non-lunghi); devono dunque essere proprietà relative. Relative a cosa? Al tempo: devo *temporalizzare* le proprietà, così dirò che ho i capelli *lunghi-in- t_1* e *corti-in- t_2* . Nel presentismo questo vale per *tutte* le proprietà, nel presentismo sono tutte relative al tempo; di conseguenza, nel presentismo non esistono proprietà intrinseche, e questo è difficilmente accettabile. Com'è possibile che proprietà contraddittorie siano intrinseche? Devo temporalizzare il soggetto, cioè devo dargli *parti temporali*: l'*io-di-oggi* ha intrinsecamente i capelli lunghi, l'*io-di-domani* ha intrinsecamente i capelli corti. Io devo quindi essere un ente quadridimensionale, esteso nel tempo, come vuole l'eternalismo.

L'argomentazione di David Lewis è stimolante, ma ha i suoi limiti. Infatti, chi è che ha i capelli lunghi o corti? *Io*, ovviamente: le proprietà sono possedute dagli enti nel loro complesso e non dalle loro parti, l'affermazione «ho i capelli lunghi» è vera per me e non solo per una mia parte. Senza dubbio le parti temporali di una persona non sono a loro volta persone (altrimenti dovrei dire di avere dentro di me un'infinità di persone diverse); sono invece astrazioni, le parti temporali sono fuse in un unico ente reale che io divido artificialmente. Ma se non ho vere parti temporali, e quando dico «io» mi riferisco a tutta la mia estensione temporale in un colpo solo, allora il fatto che prima abbia i capelli lunghi e poi corti continua a costituire una contraddizione, quindi sono comunque costretto a temporalizzare le proprietà rendendole estrinseche.

Se invece accetto di avere parti temporali, e che esistono sia parti passate che future, devo accettare che il passato e il futuro esistono tanto quanto il presente. Questo, però, non spiega perché il presente è *dominante* rispetto agli altri tempi: io non vedo direttamente il passato né il futuro, io vedo solo il presente; gli enti del passato e del futuro non hanno potere causale, sono *astratti*, un muro esistente solo nel passato o nel futuro non può fermare la mia corsa, solo un muro presente può farlo, solo il presente è *concreto*. L'eternalismo deve rendere merito di ciò, quindi deve dare a passato e futuro uno status diverso da quello del presente; ma se lo fa, diventa equivalente al presentismo.

Inoltre, se è vero che ho parti temporali sia passate che future, allora il controfattuale «io sarei potuto morire a tre anni» deve essere falso: infatti «io» deve riferirsi a tutta la mia estensione temporale, l'io che sta scrivendo queste righe è un serpentone lungo almeno ventisette anni; ma un individuo morto a tre anni avrà molte meno parti temporali di me, e dunque quello non posso essere io. Epperò, il controfattuale pare vero, e se io davvero sarei potuto essere diverso da ciò che sono allora come minimo non posseggo già parti temporali future, ma questo l'eternalismo non può accettarlo.

Considero ora la proposizione espressa da «Massimino Trace era più alto di Napoleone». Si tratta senza dubbio di un'affermazione vera: Napoleone aveva un'altezza normale, mentre Massimino Trace era un gigante. Ma se è vera, allora corrisponde a dei fatti, e i fatti sono ciò che esiste *ora*. Infatti per l'eternalismo, essendo il passato esistente tanto quanto il presente, Massimino e Napoleone esistono, e quindi la proposizione è vera in senso pieno (un vantaggio rispetto al presentismo, per il quale invece deve essere falsa). Ma cosa dire di una proposizione al futuro come «domani pioverà»? Il passato è unico, fisso, inerte e immutabile; il futuro, invece, è indeterminato, ci sono più futuri possibili, e non possono esistere tutti. Certo, «domani pioverà» è o vera o falsa perché domani o pioverà o non pioverà; ma quando pronuncio la frase, non esiste alcun fatto in merito, potrebbe darsi un caso come l'altro. Le proposizioni al futuro, allora, dovrebbero essere indeterminate o tutte false, ma l'eternalismo non riesce a rendere merito di questo.

Se voglio salvare l'eternalismo, non posso continuare a sostenere che passato, presente e futuro siano tutti e tre ugualmente esistenti: devo rinunciare ad uno di questi. L'esistenza del presente è indiscutibile. Potrei sostenere che esistono solo presente e futuro, e non il passato: dipingerei un universo nichilista, un fiammifero che si consuma, lo scorrere del tempo sarebbe il disintegrarsi di un'appendice, una disintegrazione che avanza fino a cancellare tutta l'esistenza. Ma questa teoria è difficilmente difendibile per motivi già esposti. Devo sostenere, invece, il *growing block universe*, l'idea che il passato e il presente esistono, ma non il futuro. In tale teoria, il passato è inerte e statico, mentre il presente è in continua evoluzione; io sono un serpentone temporale in crescita, ho una coda che porta alla mia origine e una testa che si fa strada attraverso i futuri possibili, cosicché sono nato e non morirò, perché continuerò ad esistere eternamente nel passato.

Epperò anche l'idea dell'esistenza del passato pone dei problemi. Io, come il grande logico Prior dal quale ho preso tale argomentazione, soffro di emicrania, un'emicrania che senza medicinali mi andrebbe avanti per giorni e giorni. Quando mi passa, dico «grazie a Dio è finito». È finito? Se il passato è reale ed eterno, le cose non finiscono, il mio mal di testa continua ad esistere nel passato. Dovrei dire allora «grazie a Dio è nel passato»: ma dovrebbe essere questa una cosa della quale rallegrarmi, visto che il mio mal di testa continua ad essere reale?

Altro esempio. In un certo momento della sua vita Aristotele avrà pensato «ora sto pensando». Se il passato è esistente, allora posso dire che esiste Aristotele e che sta pensando «ora sto pensando». Il suo pensiero è vero o falso? Se passato e presente esistono contemporaneamente e allo stesso

modo, allora la parola «ora» si riferisce al *vero* presente, quello nel quale io vivo, e la sua frase è falsa. Ma questa mossa è illegittima: Aristotele, infatti, non può sapere di vivere nel passato e non nel presente; e io sono nella stessa situazione, io non posso essere sicuro di non essere uno di quegli enti che esistono solo nel passato. Ma questa ipotesi è troppo assurda per essere presa sul serio: è più facile negare direttamente l'esistenza del passato.

Il presentismo non rende merito di tutti i fatti

Se passato e futuro non esistono in senso pieno, cosa dire del solo presente?

Il principale argomento contro il presentismo poggia sulla riducibilità del tempo al cambiamento. Se non vi fosse il tempo, non potrebbe esservi alcun cambiamento: come nei film di fantascienza, se sono seduto in una certa posizione e il tempo si ferma, non potrò più cambiarla finché non torna a scorrere. Ma se non vi fosse il cambiamento, potrebbe esservi comunque il tempo? Se l'universo fosse del tutto immobile, se fosse una realtà statica sempre identica a sé stessa in ogni sua parte, potrei dire comunque che il tempo sta scorrendo? L'intuizione comune è che non vi sarebbe alcun motivo per postularlo. Stando così le cose, il tempo potrebbe essere solo un modo di concettualizzare il cambiamento, che dipende a sua volta dalle relazioni tra gli enti: e se è così, il tempo non è oggettivo e assoluto come vorrebbe il presentismo.

Il filosofo Shoemaker solleva un'obiezione. Immagino vi sia un universo diviso in due. La metà sinistra è tale che, ogni due anni, va in *freeze* per un'ora, per un'ora in quella metà non si riscontra più alcun cambiamento; la metà destra, invece, è tale da andare in *freeze* per un'ora ogni tre anni. Se osservo questo universo per due, tre, quattro anni, vedrò in *freeze* sempre e solo una delle due parti, e quindi saprò che il tempo continua a scorrere in virtù dei cambiamenti nell'altra; ma proprio per questo, al sesto anno, saprò che il tempo scorre anche se tutto l'universo rimane immobile. Se ho la possibilità di studiare le leggi che dominano quell'universo (per esempio, se ogni *freeze* è preceduto da un bagliore particolare) posso calcolare che ogni sei anni c'è un'ora di *tempo vuoto*, tempo senza cambiamento che cionondimeno esiste.

Questa argomentazione mostra che l'idea di un tempo senza cambiamento non è logicamente aberrante: ipotizzare che ogni singolo secondo questo universo vada in *freeze* per un miliardo di anni è sempre epistemologicamente ingiustificato, ma non è un'impossibilità logica. Epperò: in che senso il tempo continuerebbe a scorrere quando nulla cambia? È come se stessi immaginando una quantità, la posizione di una lancetta, una coordinata temporale che, fuori dall'universo, continua a cambiare: io posso concepire il tempo scorrere in un universo senza cambiamento solo se immagino qualcosa di esterno all'universo che continua a cambiare. Ma se è così, l'argomentazione di Shoemaker presuppone ciò che vuole dimostrare: dicendo che in metà universo «nulla cambia per un'ora» sta presupponendo l'esistenza del tempo e la sua possibilità di scorrere e quindi cambiare al di fuori dell'universo; e così non può dimostrare la possibilità metafisica di un tempo vuoto.

Un altro problema del presentismo riguarda appunto lo scorrere del tempo. Se esiste solo il presente, e non esiste il futuro, be', verso cosa sta scorrendo il tempo? A che velocità? Avanza di sessanta secondi ogni minuto? Posso rispondere che queste sono domande senza senso: il tempo non scorre, bensì ciò che è presente cambia, e non c'è altro da aggiungere. Ma se gli enti presenti *cambiano*, posso dire anche *erano altro*, forse anche che *saranno altro*, e quindi che hanno una

estensione temporale, che ci sono stati di cose passati (o futuri) più o meno lontani dal presente. Queste sono intuizioni forti che però il presentismo difficilmente accetta.

Come nel caso del naturalismo, anche il presentismo è difficilmente definibile. Cosa vuol dire «esiste solo il presente»? Se rispondo «che esiste solo *ciò che esiste ora*» sto offrendo un'analisi circolare insoddisfacente: non posso spiegare «presente» tramite «esistente ora», non è affatto più chiaro. Se rispondo invece «che è presente tutto ciò che esiste» finisco per dire qualcosa che l'eternalista accetta volentieri: poiché il passato e il futuro esistono, il passato e il futuro sono presenti, quindi il presentismo è vero perché equivalente all'eternalismo... ma questo non è certo ciò che il presentista desidera. Il presentismo così espresso è, non una teoria del tempo, ma dell'esistenza, si limita a far equivalere esistenza e presenza.

In effetti, pare che il presentismo debba riconoscere che anche passato e futuro sono presenti. Non posso dire che il mio cane Mozart «ha il muso marrone» perché non è *più* vero, lo era anni fa, e, se voglio affermare l'esistenza solo del presente, non posso dire neppure che «il *Mozart-di-anni-fa* ha il muso marrone» perché le parti temporali richiedono l'esistenza del passato e del futuro; quindi temporalizzo la proprietà e dico «Mozart ha, ora, il muso *marrone-in-gioventù* e *bianco-in-vecchiaia*». Ma: a quali fatti cercano di riferirsi queste proprietà? Sempre e comunque a fatti nel passato (la gioventù) o nel futuro (la vecchiaia); e se non esistono, il riferimento fallisce, qualsiasi affermazione contenente proprietà temporalizzate continua a risultare falsa.

Posso rispondere che le proposizioni al passato e al futuro sono sì frecce puntate verso un altro tempo, ma non lo fanno con l'intento di riferirsi ad un fatto *esistente*, bensì sono rese vere dal fatto che *è esistito* o *esisterà* in quel tempo, la frase è vera non perché *si riferisce* ma perché *si sarebbe riferita* o *si riferirà* ad un certo fatto. Se io lanciai una moneta e dico «uscirà testa» non mi riferisco ad un fatto esistente nel momento in cui pronuncio la frase, ma un fatto che *esisterà* quando la moneta toccherà il suolo. Prior era disposto ad accettare che tutte le frasi al futuro fossero false, ma questo appare troppo controintuitivo; l'interpretazione che ho proposto è una possibile soluzione, anche plausibile, ma poiché richiede di modificare una delle basi della filosofia, la teoria corrispondentista della verità, per crearne una versione temporalizzata, se è possibile è meglio agire prima su altri punti.

Detto tra parentesi: l'argomentazione del buco, ovvero «cosa ne direbbe Einstein?»

L'influenza di Einstein sulla concezione dello spaziotempo è difficilmente sovrastimabile. È sua l'idea alla base di questa argomentazione, detta *del buco*. Prima di esporla, però, ci sono degli aspetti che devo tenere in considerazione.

Einstein ha creato la base di questa argomentazione (poi rivista e sviluppata da diversi filosofi) per disperazione: aveva in mente una teoria fisica con delle caratteristiche formali particolari, ma non riusciva a crearla, così ha immaginato questa argomentazione del buco per spiegare perché una teoria del genere fosse anche solo in linea teorica irrealizzabile. Tale teoria, però, era quella della relatività, che è riuscito a realizzare; quindi l'argomentazione del buco sembra dover avere qualche errore alle sue fondamenta, anche se quali siano non è chiaro.

D'altro canto, la teoria della relatività è a sua volta sbagliata: pur essendo la miglior descrizione disponibile per certi fenomeni, non è la descrizione perfetta, è una descrizione incompatibile con la fisica quantistica, e il motore principale dietro la ricerca in questo campo è proprio il voler

sviluppare una nuova teoria quantistica della gravità che possa prendere il posto della relatività. Stando così le cose, è illegittimo dare per vera la relatività e usarla per trarre conclusioni sulla natura dello spaziotempo; ma è anche illegittimo partire dal presupposto che sia impossibile rivedere l'argomentazione del buco in modo tale da renderla ancora valida in relazione ad una diversa teoria.

Ordunque, ecco l'argomentazione.

La relatività crea un modello dell'universo usando tre elementi: M, g, T .

L'elemento M è la *varietà*, un oggetto matematico immaginabile come una serie di punti, o un reticolo, o un sistema di coordinate. Ha quattro dimensioni, e quindi quattro coordinate, che permettono di *etichettare* gli eventi dell'universo in base alla loro posizione.

La varietà però è un oggetto semplice, non ha struttura; ci sono dei punti, ma non posso ancora dire quanto siano distanti tra loro, né se una linea che passa attraverso vari punti sia una retta o meno; posso assegnare quattro coordinate, ma non posso ancora distinguere la coordinata temporale dalle tre spaziali, non posso dire quali eventi sono simultanei, quali sono passato o futuro rispetto ad un altro. Per poter compiere queste operazioni geometriche devo aggiungere l'elemento g , cioè la *metrica*; così, etichettando gli eventi con le quattro coordinate indico la loro posizione nello spazio e nel tempo. Per esempio, l'evento della mia nascita sarà etichettato con le coordinate spaziali che individuano l'ospedale di Motta di Livenza e la coordinata temporale corrispondente al 7 giugno 1993; la mia vita può essere così rappresentata come una linea che si sposta nel tempo e nello spazio.

L'ultimo elemento, T , indica ciò che può esistere all'interno dello spaziotempo, cioè *energia* e *materia*.

Con Putnam ho mostrato come *trasformare* una teoria per creare più interpretazioni, come cambiare il collegamento tra le proposizioni e gli oggetti del mondo senza cambiare i valori di verità. Ora posso compiere un'operazione simile creando due modelli collegati da una funzione detta *diffeomorfismo*, due modelli che rappresentano due modi di *proiettare* g e T sulla varietà M , cioè due modi di applicare sistemi di coordinate all'universo. Posto il primo modello M, g, T creo il secondo modello M^*, g^*, T^* che contiene, rispetto al primo, un *buco*: cioè il secondo modello usa un diverso sistema di coordinate M^* tale che g^* e T^* sono identici a g e T in ogni punto tranne che per un certo insieme di coordinate che delimitano una zona di spazio o tempo (il buco appunto). Posso immaginare, per esempio, un universo M^*, g^*, T^* identico all'universo M, g, T , con la sola differenza che esiste un certo tempo t tale che, prima di t , g^* e T^* sono identici a g e T , e dopo t sono diversi, creando così due modelli che descrivono una identica storia dell'universo fino a oggi, ma che prevedono diversi eventi futuri.

Questi due modelli hanno aspetti *varianti* e aspetti *invarianti*. Mi faccio delle domande come «che evento c'è alle coordinate $0, 0, 0, 0$?» oppure «posta la linea della mia vita, il punto di coordinate $1, 2, 3, 4$ appartiene alla linea?» Questi sono aspetti varianti: cambiando il sistema di coordinate per creare il secondo modello cambio anche la distribuzione degli eventi rispetto ai punti all'interno del buco, e quindi cambio le risposte che il nuovo modello dà a queste domande. Ora mi chiedo invece: «quanto distano temporalmente questi due eventi, cioè due punti, uno dall'altro?», «qual è la massa di questo corpo, e da quante cellule è composto?», «lo spaziotempo è piatto o curvo?», «la macchina sta accelerando?» Questi sono aspetti invarianti: il cambiare il sistema di coordinate è irrilevante in questi casi, quindi i due modelli danno le stesse risposte a tutte queste domande.

Il punto centrale dell'argomentazione è questo: che gli *osservabili*, essendo indipendenti dal sistema di coordinate scelto, sono invarianti. Cioè, i due modelli rispondono davvero in modo diverso a certe domande, ma rispondono in modo identico a tutto ciò che è empiricamente controllabile. I due modelli corrispondono a due teorie che sono sì diverse, ma che descrivono una identica *apparenza*: sia che il mondo funzioni in accordo con la prima teoria, sia che funzioni in accordo con la seconda, comunque apparirebbe proprio così come appare. Per decidere quale teoria sia migliore dovrei trovare delle differenze nelle loro previsioni, ma tutti i loro elementi varianti sono inosservabili, e quindi non sottoponibili a test empirico.

La capacità di prevedere eventi è insufficiente a trarre conclusioni sulla natura delle cose: esiste una realtà oggettiva, ma la capacità di una teoria di fare previsioni sugli eventi osservabili, sul modo in cui il mondo appare, non basta a renderla una rappresentazione perfetta di tale realtà. Questo vuol dire che la scienza è insufficiente a dire qualcosa d'interesse filosofico sulla natura dello spaziotempo; con l'aiuto della riflessione filosofica, però, proprio questo stesso fatto può spingere a favorire una certa teoria.

Se lo spaziotempo è una realtà oggettiva, cioè se M e g si riferiscono a qualcosa di reale e non sono solo astrazioni matematiche, allora, poiché i due modelli rappresentano due diversi spaziotempi, o è vero uno, o è vero l'altro. Ma questo è un caso indesiderabile: vorrebbe dire che esiste un insieme infinito di teorie, cioè tutte le interpretazioni che posso sviluppare cambiando il sistema di coordinate, che rappresentano sistemi fisici diversi e incompatibili, ma che io sono impossibilitato a distinguere tramite qualsivoglia test.

Devo chiedermi invece: *perché* i due modelli sono empiricamente equivalenti, *perché* fanno identiche previsioni per tutti i test? In matematica uno stesso ente può essere rappresentato in più modi. Per esempio, l'unità io la posso rappresentare come «1», ma anche come « $0, \bar{9}$ », sono due modi diversi di concepire la stessa realtà (se $1 = 0, \bar{9}$ allora $10 = 9, \bar{9}$, quindi $10 - 1 = 9, \bar{9} - 0, \bar{9}$, e infatti $9 = 9$). Forse può dirsi lo stesso per i modelli M, g, T ed M^*, g^*, T^* : sono due rappresentazioni della stessa realtà, non ho motivo di credere che rappresentino realtà diverse quando sono identiche sotto tutti gli aspetti osservabili. Se è così, allora il fatto che una teoria sia trasformabile nel modo visto è accettabile perché cambiando il sistema di coordinate non cambio la realtà rappresentata; il che vuol dire però che M e g non si riferiscono ad un ente oggettivo. In altre parole, l'ente al quale si riferiscono è più coerente con quello immaginato dall'eternalismo, che dal presentismo.

È chiaro che tale conclusione non è una necessità logica: potrebbe essere vero che i due modelli rappresentano due realtà distinte, e poiché esiste una sola realtà sarebbe vero che un certo punto o appartiene o non appartiene ad una retta, che ad una certa coordinata c'è o non c'è un certo evento. Questi aspetti varianti, però, sono incontrollabili, non posso sapere se sia effettivamente così. E poiché questo mio non poterlo sapere implica indesiderabilmente anche l'impossibilità di sapere com'è davvero strutturata la realtà e come evolve nel tempo, ho un motivo pragmatico per credere che non sia così, e quindi un motivo pragmatico per preferire l'eternalismo al presentismo, per preferire la teoria che più facilmente accetta uno spaziotempo convenzionale e relativo anziché oggettivo e assoluto.

Questo può sembrare un po' poco, però, per giustificare la credenza nell'eternalismo. Negli anni sono state sviluppate ampie e complesse analisi dell'argomentazione, e la forza delle sue conclusioni dipende da quali delle componenti M, g, T si considerano essere lo spaziotempo. Io sono partito dal presupposto che lo spaziotempo sia composto come minimo da M e g , che chi

sostiene l'esistenza oggettiva dello spaziotempo sostenga l'esistenza oggettiva di quanto indicato da M e g . Ciò pare sbagliato: M e g combinati si riferiscono ad un ente troppo complesso e specifico per essere uno spaziotempo, quindi l'argomentazione parte da una premessa errata. Di contro, però, neanche M da solo può essere lo spaziotempo, perché indicherebbe un ente troppo semplice.

Quindi è necessaria una diversa teoria del tempo

Il presentismo e l'endurantismo sono le teorie più intuitive, ma nel corso del secolo scorso eternalismo e perdurantismo sono diventate popolari a causa dell'autorità assegnata dal naturalismo alle argomentazioni della fisica. Oggi però il consenso tra i filosofi è che nessuna di queste due coppie rappresenti una teoria adeguata, hanno entrambe grandi punti deboli ed è necessario sviluppare posizioni nuove.

Una prima mossa allora è provare a scambiare un elemento delle coppie. Per esempio posso sostenere insieme eternalismo ed endurantismo: passato e futuro esistono, ma gli oggetti sono pienamente presenti in ogni istante senza avere parti temporali. L'altra possibilità, però, è più attraente: infatti il presentismo continua ad essere la posizione più vicina al senso comune, mentre il perdurantismo ha innegabili vantaggi nella gestione dell'identità. Come mettere d'accordo queste due posizioni, l'esistenza del solo presente con l'esistenza di enti estesi nel tempo?

L'idea è questa: esiste solo il presente, l'ente è presente, ma la parte presente dell'ente non lo esaurisce, l'ente è più ampio del tempo perché ha sia una parte temporale (presente) sia una atemporale (fuori dal tempo, né passata né futura).

È una fantasia bizzarra. Di norma il presentista nega non solo l'esistenza di passato e futuro, ma anche la possibilità di enti atemporalmente. Dal canto mio, mi rendo conto di non avere le risorse intellettuali necessarie a poter comprendere cosa potrebbe essere un'esistenza atemporale, sarebbe come capire cosa vuol dire *essere* una legge fisica (ammesso che le leggi fisiche creino il tempo e non esistano *in* esso). Il fatto che, in questo caso, io stia immaginando un ente che ha una parte temporale e una atemporale, oltre a rimanere incoerente con l'inesistenza di enti atemporalmente, rende la situazione ancora più radicalmente incomprensibile.

Devo fare un ulteriore sforzo per cercare di rendere la posizione intelligibile:

- esistono enti semplici (senza parti) ed enti complessi (divisibili in parti);
- per gli enti semplici, che non possono cambiare nel tempo, è vero il presentismo endurantista, cioè sono pienamente esistenti in ogni istante presente;
- per gli enti complessi, essi hanno sia parti per le quali è vero il presentismo endurantista (parti pienamente esistenti nel presente), sia *parti** per le quali è vero l'eternalismo perdurantista (parti atemporalmente, cioè composte a loro volta da parti temporali ma senza alcun ordinamento oggettivo);
- quando una parte presente cessa di essere tale continua la sua esistenza come *parte**.

L'idea che un ente abbia parti perduranti presenti e parti enduranti eterne può avere qualche riflesso in teologia, se accetto che il corpo temporale sia tutt'uno con un'anima atemporale; può essere anche un modo di spiegare come sia possibile avere conoscenza di fatti astratti non-naturali, non-interagenti in quanto esterni allo spaziotempo, come potrebbero essere i fatti etici (cioè:

l'uomo può conoscerli perché ha a sua volta una parte esterna allo spaziotempo), anche se si tratta di una spiegazione poco convincente.

Alla fine, la teoria più facilmente accettabile è quella di un presentismo accompagnato da una teoria della verità temporalizzata nella quale ogni affermazione si riferisce ad un certo fatto *in uno specifico tempo* (dato sistematicamente per implicito). Ciò nonostante, l'idea di temporalizzare i *truthmaker* potrebbe non essere altrettanto facile da accettare, e se non lo è, tra filosofia del tempo e teoria della verità è la prima a dover cambiare.

LETTURE PER APPROFONDIRE

blabla bla

blablabla

blablabla

Per comprendere la moderna concezione dello spaziotempo bisogna capire la teoria della relatività, il che richiede conoscenze matematiche piuttosto avanzate. Fortunatamente Clement Durell ha pensato bene di scrivere "La Relatività con le Quattro Operazioni", un testo nel quale si spiega la teoria della relatività senza rinunciare alla matematica, ma semplificando fino a ottenere equazioni comprensibile a chiunque abbia fatto una qualsiasi scuola superiore. Vi sono poi due libri popolari di Einstein, "Pensieri, Idee, Opinioni" e "Come io Vedo il Mondo - La Teoria della Relatività", l'ultimo dei quali contiene tre paper di notevole rigore matematico. Abbastanza complesso, e strano, ma anche più aggiornato, "La Natura dello Spazio e del Tempo" di Hawking e Penrose; testo strano, dico, perché affianca esposizioni verbali divulgative a espressioni matematiche che richiedono conoscenze universitarie. Più divulgativi e semplici "Sette Brevi Lezioni di Fisica" e "L'Ordine del Tempo" di Carlo Rovelli, nei quali pure si trovano idee rilevanti. Sempre divulgativo, ma richiedente molte più energie, l'incredibile "La Strada che Porta alla Realtà" di Penrose, un libro scritto magnificamente, pieno di inventiva artistica, ma lungo più di mille pagine, che porta da zero alla teoria dei quanti introducendo man mano tutti gli strumenti matematici essenziali. Per approfondire la fisica moderna dopo Einstein (che ha ulteriormente modificato la propria comprensione dello spaziotempo) si può iniziare con "Sei Pezzi Facili" e "Sei Pezzi Meno Facili" di Feynman, libri scritti con notevole ironia, per poi passare a "I Principi Fisici della Teoria dei Quanti" di Heisenberg, che richiede invece una ben più profonda comprensione della matematica. Il più importante filosofo del tempo in filosofia analitica è un non-analitico, J. M. E. McTaggart, con il paper "The Unreality of Time" in cui introduce la distinzione tra serie A e serie B. Per la concezione relazionista dello spaziotempo, legata all'eternalismo e al perdurantismo, vi è da prendere in forte considerazione il lavoro di Ernst Mach, fisico, fisiologo e psicologo tra i più originali di sempre (per esempio si veda il suo "I Principii della Meccanica"). Per l'argomentazione in merito di Putnam, anche se questa è esposta nel suo paper "Time and Physical Geometry", è più facile trovare il libro di Penrose "La Mente Nuova dell'Imperatore" che ne espone una sua versione. Sull'argomentazione "del buco" più di tutti ha scritto John Earman, che ha pubblicato lungo tutta la sua carriera un gran numero di ricerche sullo spaziotempo ("World Enough and Spacetime", "In the Beginning, at the End, and All in Between"); si veda però anche "Foundations of Space-Time Theories" di Michael Friedman e, per una prospettiva più generale, "Time, Tense, and Causation" di Michael Tooley. Per quel che riguarda la logica modale temporale, rilevante per

la filosofia dello spaziotempo, il più grande filosofo è Prior, autore di numerosi lavori come "Time and Modality", "Past, Present and Future" e "Papers on Time and Tense".

Commentatore di McTaggart, C. D. Broad implica dal suo lavoro una filosofia dell'identità ("Personal Identity and Survival"). Molto diretto anche C. J. F. Williams con titoli come "What is Identity e "Being, Truth and Identity". Il perdurantismo è sostenuto soprattutto all'interno della filosofia australiana, che è una filosofia analitica molto coraggiosa e stravagante nel tentare l'analisi di posizioni anche estreme. Penso soprattutto a David Malet Armstrong (in italiano sono disponibili "Ritorno alla Metafisica" e "Cos'è la Metafisica") e a David Lewis, inglese ma associato alla filosofia australiana (il libro in merito è sempre "Sulla Pluralità dei Mondi"). È curioso che sia D. M. Armsfrong che Lewis, pur essendo atei, pensassero alla possibilità di una sopravvivenza personale alla morte: così Lewis fa nel suo ultimissimo paper, "How Many Lives Has Schrodinger's Cat?", ironicamente pubblicato postumo. Si consideri anche "Four-Dimensionalism" di Ted Sider. Il miglior libro sull'identità personale rimane "Ragioni e Persone" di Parfit, nel quale si trovano molti esempi citati; molto importanti pure continuano a essere "Saggi sull'Intelligenza Umana" di Locke e "Ricerca sull'Intelletto Umano" di Hume, nonché "Nome e Necessità" di Kripke e "Sulla Pluralità dei Mondi" di Lewis. L'esempio relativo allo scambio di mente, invece, si trova in "The Self and the Future" e in "Body Continuity and Personal Identity" di Williams. Sull'interpretazione vaga e relativa dell'identità si vedano i paper di Chisholm ("The Loose and Popular and the Strict and Philosophical Senses of Identity") e Geach ("Identity", "Ontological Relativity and Relative Identity"). Nel campo della mereologia il più grande e forse anche miglior filosofo italiano è Achille Varzi. Molti dei suoi lavori migliori sono fatti in collaborazione con Robert Casati, anche lui uno dei filosofi italiani più interessanti nonché autore di storie appassionanti: si vedano per esempio "Buchi e altre Superficialità" e "Parts and Places". Culturalmente, la mereologia può essere considerata come appartenente a quel clima che ha prodotto anche la cibernetica e la scienza dei sistemi, discendente dal lavoro interdisciplinare di un certo numero di geni uno più stravagante dell'altro (Babbage, Heaviside, Geddes): questa connessione, a mio parere, è evidente nel lavoro di George Spencer-Brown "Laws of Form", una sorta di logica che permette di svolgere calcoli dividendo la pagina in parti.

In merito alla necessità dell'identità, si veda "Identity of Individuals in a Strict Functional Calculus of Second Order" di Marcus, "Can There Be Vague Objects?" di Evans e "Constitution and Kind Membership" di Rea. L'esempio del "tempo vuoto", tempo senza cambiamento, si trova nel paper "Time Without Change" di Shoemaker.

Metafisica: filosofia del libero arbitrio

Sesto ragionamento: una difesa del libertarismo

In questo capitolo difendo una teoria libertarista facendo uso di argomentazioni spesso trascurate a livello accademico. Lo schema è il seguente:

- P1** non ci sono prove a favore del determinismo;
- P2** la libertà è presupposta dalla pratica scientifica;
- P3** la libertà è necessaria alla conoscenza;
- C1** quindi il libertarismo deve essere vero;
- P4** l'esistenza di più futuri possibili è necessaria alla libertà;
- P5** l'originazione è necessaria alla libertà;
- P6** il compatibilismo non può accettare entrambi i principi;
- C2** quindi non può esserci compatibilità tra libertarismo e determinismo.

Le tre desi: libertarismo, determinismo, compatibilismo

Gli scrittori che ammiro davvero si contano sulle dita di una mano: Goethe, Proust, Joyce, d'Annunzio; e l'ultimo è Tolstoj. Ho un'ossessione per la sua vita eroica. Non concordo con l'assioma che lo vede come il miglior scrittore del suo secolo ma il peggior filosofo: il suo pensiero è radicale, estremo, ma lo apprezzo più di quello di certi *scrittori-filosofi* come Dostoevskij o Sartre.

Un'estate di tanti anni fa ho letto il volume unico di Guerra e Pace in spiaggia: non certo il libro più agile da maneggiare stando sullo sdraio, né il più coerente con la sabbia bollente tutt'attorno. Tolstoj alterna parti narrative e fulminei saggi sulla Storia. In uno di tali capitoli saggistici riflette sulla libertà, e dice: guarda un po', decido di alzare la mano destra, e la alzo, e se non voglio la tengo giù. C'è un contrasto tra l'esperienza di vita e la vita com'è descritta dagli storici: io ho l'impressione di fare ciò che faccio perché così decido, sono libero di fare come di non fare a seconda della mia volontà; gli storici, invece, vedono sempre motivazioni dietro gli eventi, contingenze economiche ed altri fattori che causano gli eventi futuri senza lasciare spazio alla libertà umana. Infatti, come sarebbe possibile spiegare gli eventi, se tanto l'uomo agisce in modo indipendente da qualsiasi evento passato? Dovrebbe essere tutto casuale. Dunque, nella mia esperienza esistono più futuri possibili che posso scegliere liberamente, nella visione dello storico c'è un unico futuro inevitabile determinato dagli eventi passati: chi ha ragione?

La filosofia analitica offre tre risposte possibili. La prima è il *determinismo*: non c'è libertà, non c'è libero arbitrio, c'è un unico futuro possibile e io posso fare solo ciò che è determinato io faccio. La seconda è il *libertarismo*: c'è libero arbitrio, ci sono più futuri possibili e, benché io sia influenzato, nulla mi determina a scegliere una possibilità d'azione rispetto ad un'altra. Sia determinismo che libertarismo sono posizioni *incompatibiliste*, cioè sostengono l'incompatibilità di libertà e determinazione. La terza risposta è il *compatibilismo*: i concetti di libertà e determinazione sono coerenti tra loro, e posso quindi dire che è vero che sono determinato, ma è anche vero che sono libero.

Non ci sono prove a favore del determinismo 1: Battaglie navali e verità

Il concetto di *libero arbitrio* come *libertà di fare altrimenti* nasce con la filosofia moderna, ed è un errore leggere gli scritti dei filosofi antichi come se avessero già accesso a questo stesso concetto. Per esempio, nella filosofia medievale si parla di *libertà*, ma intendendola solo come *libertà di fare il bene o il male*, un concetto più limitato di quello attuale. Ciò nonostante, l'idea intuitiva della determinatezza, l'idea di essere guidati da un destino che si realizzerà infallibilmente, è presente in tutta la tradizione occidentale sin dai primi racconti mitologici.

Questa linea di ragionamento si trova già nei filosofi presocratici: una proposizione al futuro come quella espressa da «domani vi sarà una battaglia navale» è, già nel momento in cui la pronuncio, o vera o falsa; ma se è vera già oggi, allora è determinato che domani vi sarà una battaglia navale, e se il futuro è determinato non può esserci libertà.

La conclusione appare frettolosa: anche se quella proposizione è vera, non è *necessariamente vera*, e se non lo è non c'è determinatezza. Così, se io, lanciando una moneta ideale, dico «uscirà testa», dico una cosa vera se cadrà di testa, ma il risultato del lancio è davvero casuale, è davvero indeterminato, quindi la mia frase è solo *contingentemente vera*; similmente, se la verità di «domani vi sarà una battaglia navale» è contingente, il futuro può rimanere indeterminato.

Ma è davvero così? In Aristotele si trova una controargomentazione. Sia che io accetti la contingenza e indeterminatezza del futuro, sia che la rifiuti, comunque posso accettare la necessità del passato: esso è statico e immutabile, il passato non può più essere diverso da com'è, non può essere modificato, quindi la verità di «due giorni fa ho mangiato lenticchie» è oggi necessaria. Dunque:

- per ipotesi, oggi è contingentemente vero che «domani vi sarà una battaglia navale», perché vi sarà ma poteva anche non esservi;
- se è vero oggi, allora lo era anche diecimila anni fa;
- poiché il passato è necessario, non è possibile che «diecimila anni fa era vero che domani vi sarà una battaglia navale» fosse falsa, è necessariamente vera;
- ma se fosse possibile che «domani vi sarà una battaglia navale» sia falsa, sarebbe possibile anche per questa sua variante al passato;
- poiché non è così, anche «domani vi sarà una battaglia navale» deve essere necessariamente vera, quindi c'è un unico futuro possibile, quindi il futuro è determinato e non può esservi libero arbitrio.

Come già accennato, però, diversi filosofi non credono che il passato sia necessario: se è vero che il concetto di *retrocasualità* è coerente, se è logicamente possibile che un evento futuro sia la causa di

un evento passato, allora, se il futuro è contingente, anche il passato può esserlo. Per questo l'argomentazione commette una *fallacia modale*: dal fatto che, in un certo tempo, sia vero che vi sarà una battaglia navale, cerca di dedurre che è *necessariamente* vero, ma si tratta di una deduzione illegittima; è vero che al fulmine segue il tuono, ma da ciò non posso dedurre che al fulmine segua *necessariamente* il tuono, è necessario che dopo il fulmine *o vi sarà o non vi sarà* il tuono, ma nessuna delle due possibilità è necessaria.

In linea con certe ipotesi viste in filosofia del tempo, posso invece immaginare una verità *retroattiva*. Se un omosessuale viene condannato dieci anni fa, e ora l'omosessualità viene depenalizzata, non solo quella persona è ora innocente, ma riconosco che lo è sempre stata, che non è mai esistito un simile reato neanche quando era in carcere. Similmente, quando io dico «domani vi sarà una battaglia navale», se dico qualcosa di vero lo dico in virtù del fatto che *vi sarà* domani, ed è questo fatto che, quando è presente, rende retroattivamente vere le frasi al futuro pronunciate nel passato. Questo non rende la verità di tali frasi necessaria, quindi non implica un futuro determinato.

Non ci sono prove a favore del determinismo 2: il demone di Laplace e la scienza

Uno dei pilastri della filosofia di Leibniz è il *principio di ragion sufficiente*, cioè: ogni cosa accade per una ragione, nulla esiste senza una ragione. Se decido di mangiare, è perché ho fame e così funziona il mio stomaco; se lavoro, è perché così funziona la società: ogni evento ha una ragione, cioè una causa.

Questo principio intuitivo sta alla base della scienza, che si stava sviluppando all'epoca di Leibniz: la scienza è appunto un metodo per spiegare il mondo tramite la sua struttura causale. E come tale principio tende a negare la mia libertà (in quanto ogni mio comportamento ha una causa a me esterna che lo precede), così i deterministi hanno subito sfruttato la scienza nelle loro argomentazioni. Newton ha mostrato che il movimento dei pianeti è spiegabile tramite semplici formule, non sono liberi di muoversi, seguono una legge matematica e non possono fare altro. Se questo vale per i pianeti, probabilmente vale anche per il mio corpo, le mie cellule, i miei atomi: tutto in me sta obbedendo alle leggi fisiche, ogni mio atomo si muove in un certo modo e non può muoversi in alcun altro, quindi non ho alcuna base per ritenermi libero.

Uno dei primi esperimenti mentali sulle implicazioni della visione scientifica del mondo sul libero arbitrio è quello del *demone di Laplace*. Laplace fu un grande matematico, fisico, astronomo e ingegnere, la magnitudine della sua influenza sul pensiero scientifico è paragonabile a quella di Faraday o Huygens, cioè dei più grandi. Laplace dà il suo nome a questo esperimento, ma non è l'unico ad averlo formulato: lo si trova anche in Ruggero Giuseppe Boscovich, pure lui grande matematico, fisico, astronomo, teologo gesuita, filosofo, diplomatico, poeta e molte altre cose, un vero polimata. L'idea è quella di un'intelligenza, molto più eminente di quella umana, in grado di conoscere la proprietà di ogni singolo atomo presente nell'universo: per le conoscenze scientifiche dell'epoca, un simile essere doveva essere in grado di calcolare ogni evento accaduto in passato fino alla creazione dell'universo, e anche ogni evento futuro fino alla sua fine, descrivendo così un universo assolutamente determinato.

La contemporanea visione scientifica del mondo è però molto diversa da quella di Laplace. Il Novecento ha visto lo sviluppo della fisica quantistica, che descrive un livello della realtà più fondamentale di quello della fisica classica, e mentre quest'ultima descriveva un mondo determinato, la fisica quantistica pare descrivere un mondo intrinsecamente probabilistico e

genuinamente indeterminato, con eventi, come il decadimento degli atomi, che paiono davvero casuali. Invero, esistono più interpretazioni della fisica quantistica, alcune che accettano un mondo indeterminato, e altre che sono del tutto deterministiche; epperò gli sviluppi recenti della fisica sono comunque sufficienti a spiegare perché il demone di Laplace è impossibile.

In primo luogo, il demone di Laplace si basa sul presupposto che le leggi fisiche siano tutte *reversibili*. Se registro su videocassetta una partita di biliardo, la guardo, e poi la riavvolgo, posso vedere la partita al contrario fino a tornare alla posizione di partenza delle biglie: questo è un processo reversibile, un processo che ripetuto al contrario ricrea lo stato di partenza. La fisica moderna però riconosce l'esistenza di processi *irreversibili*: se io mescolo acqua e zucchero girando il cucchiaino verso destra, non è che girandolo verso sinistra torno a separare le due sostanze, le mescolo invece sempre di più, non è sufficiente ripetere il processo al contrario, devo invece agire sull'ambiente circostante. Stando così le cose, il demone di Laplace non può ricostruire la storia passata del mondo solo conoscendo le proprietà attuali di ogni atomo: farebbe, coi suoi calcoli, ciò che faccio io mescolando l'acqua al contrario, penso di tornare allo stato iniziale e invece creo stati diversi sempre più disordinati.

In secondo luogo, anche se così non fosse ed esistessero davvero solo processi reversibili, non avrei alcun motivo per credere in una causalità che va dal passato al futuro, un universo in cui la causalità va dal futuro al passato sarebbe perfettamente equivalente (il che non sarebbe possibile per i processi irreversibili); ma allora non avrei neanche motivo per credere in un futuro determinato dal passato. Prigogine, premio Nobel per la chimica, nota inoltre che le leggi deterministiche sono reversibili, e non prevedono quindi una linea del tempo con una direzione oggettiva, implicando così una filosofia del tempo eternalista; ma poiché i processi irreversibili esistono, esiste anche una linea del tempo che va dal passato al futuro (e non viceversa), e quindi esistono leggi non-deterministiche.

In terzo luogo, anche accettando un'interpretazione deterministica della fisica quantistica, comunque la biologia e la medicina sono ricche di casi nei quali approcci statistici si sono rivelati più utili ed efficienti di quelli deterministici, suggerendo l'esistenza di un'indeterminatezza ad un qualche livello della natura.

Il fatto che l'universo sia indeterminato, però, è insufficiente a garantire la possibilità del libero arbitrio: di sicuro se devo agire in accordo con un processo deterministico io non sono libero, ma non lo sono neanche se devo agire in accordo con un processo casuale, comunque sono costretto a fare *ciò che casualmente capita*. Quindi l'indeterminatezza può portare a rifiutare il demone di Laplace, ma non dimostra la libertà umana.

Non ci sono prove a favore del determinismo 3: Libet e la psichirurgia

Oltre alla fisica, anche la psicologia e la neurologia hanno molto da dire sul libero arbitrio: se il comportamento, che dipende dal cervello, è indeterminato, posso aspettarmi che anche l'attività del cervello lo sia, e questo può essere verificato. L'autore dell'esperimento più discusso prodotto da queste scienze è il fisiologo Benjamin Libet.

L'elettroencefalogramma mi permette di osservare l'attività elettrica del cervello. C'è un segnale, chiamato *bereitschaftspotential*, rappresentante un certo evento nella corteccia motoria, che viene osservato un attimo prima dell'inizio di un movimento. Quindi osservando

l'elettroencefalogramma e attendendo il *bereitschaftspotential* posso prevedere quando un soggetto si muoverà. Libet si chiede se questo segnale preceda anche la volontà cosciente del movimento.

Nell'esperimento di Libet i soggetti osservano lo schermo di un oscilloscopio modificato in modo tale che le variazioni nella tensione elettrica siano visualizzate come un punto che si muove in cerchio: può così essere utilizzato come un cronometro molto preciso. Viene chiesto di fare un movimento con il polso, in qualsiasi momento lo si desidera. I soggetti devono quindi memorizzare la posizione del punto nell'istante in cui hanno percepito la volontà di muoversi; un elettromiografo registra invece la posizione del punto nel momento in cui i muscoli hanno iniziato a contrarsi: calcolando la distanza tra i punti si sa quando tempo è trascorso tra un evento e l'altro.

Si nota così che: prendendo come punto zero l'istante indicato dall'elettromiografo, la volontà di muoversi è esperita duecento millisecondi prima, ma il *bereitschaftspotential* è osservato già cinquecento millisecondi prima. Conclusione: il mio cervello inizia il processo che si traduce nel movimento a livello inconscio, prima che io sia cosciente di volermi muovere, ma io non posso avere controllo volontario su ciò che è inconscio, quindi non sono libero di scegliere quando muovermi, il libero arbitrio è un'illusione.

Epperò Libet stesso dubita di queste conclusioni. Successivi esperimenti, infatti, hanno osservato che il soggetto, anche dopo l'osservazione del *bereitschaftspotential*, può porre un *veto* per interrompere l'azione, e quindi mantiene il controllo volontario sul movimento. Per di più, l'esperimento calcola quando la volontà si manifesta affidandosi a una dichiarazione in merito alla posizione del punto nel momento in cui si ricorda di aver esperito tale volontà: ma questo è un possibile difetto di design.

Lo è, in primo luogo, perché il cervello impiega alcuni millisecondi per analizzare lo stimolo e riconoscere l'immagine che ha davanti; il dover poi fare attenzione al momento in cui si percepisce la volontà (un sforzo che in circostanze normali nessuno compie) contribuisce a sovraccaricare il cervello e a rallentare. Questo significa che la posizione del punto osservata potrebbe essere successiva a quella in cui era quando si è davvero manifestata la volontà, e quindi potrebbe non esserci un *delay* tra la volontà e il *bereitschaftspotential*, o comunque potrebbe avere cause diverse da quelle indicate dalla conclusione determinista.

In secondo luogo, l'esperimento richiede ai soggetti di lasciare che la volontà di muovere il polso emerga liberamente nella coscienza, ma non ho motivo di pensare a questa emergenza come un evento libero. Nessuno crede che il libero arbitrio sia una questione di millisecondi, è invece qualcosa che riguarda ragionamenti complessi ed estesi nel tempo: io potrei decidere ora di muovere il polso dopodomani, e questo è un esercizio della mia volontà che nulla ha a che vedere con il *bereitschaftspotential* o con le sensazioni che posso esperire poco prima di attuare il movimento, né è qualcosa che può essere indicato memorizzando la posizione di un punto.

Un'altra argomentazione psiconeurologica riguarda la *psicocirurgia*, la chirurgia che punta a risolvere problemi psicologici intervenendo fisicamente sul cervello. Un'operazione psicocirurgica è la *callosotomia*, cioè il taglio del corpo calloso, la struttura che unisce i lobi cerebrali. In alcuni casi tale operazione è un efficace rimedio per l'epilessia; in altri, ha notevoli effetti indesiderati. La *sindrome della mano aliena* è uno di questi: chi viene colpito riporta che una propria mano, di solito la sinistra, agisce come avesse una sua volontà.

Ecco un caso clinico. Il paziente affetto da mano aliena riceve un sacchetto con dentro due oggetti diversi riconoscibili al tatto, un cucchiaio e un uovo; gli si chiede di prendere quest'ultimo con la

mano destra, senza guardare. Si nota allora che il paziente fruga nel sacchetto con entrambe le mani, prendendo l'uovo con la destra e il cucchiaino con la sinistra; se gli si chiede cosa sta facendo la mano sinistra, risponde che non ne ha idea. Ovviamente questa mano non è controllata da un'altra volontà, bensì pare controllata dal suo inconscio (che non è la volontà conscia di qualcun altro). Conclusione: la callosotomia, dividendo i lobi, distrugge il senso d'agenzia per le azioni controllate da uno dei due lobi, la mano si muove come sempre ma il paziente non ha senso di controllo su di essa; ma se una mano può fare tutto ciò che di norma fa anche senza il controllo volontario, questo controllo deve essere illusorio, e se questo vale per un lobo può valere anche per l'altro, quindi non ho vero controllo su alcuna azione, quindi non c'è libertà.

Anche queste conclusioni sono discutibili. Gli emisferi gestiscono processi separati: l'area a destra del campo visivo è analizzata dall'emisfero sinistro, l'area a sinistra dall'emisfero destro, l'analisi del linguaggio è svolta nella maggioranza delle persone dall'emisfero sinistro. Così, se prendo una persona che ha subito callosotomia e le faccio vedere una scimmia alla sinistra del campo visivo, le informazioni visive saranno analizzate dall'emisfero destro, ma non potranno raggiungere l'area del linguaggio a sinistra, di modo che, pur vedendo la scimmia, non sarà in grado di dire cosa sta vedendo. Al che, è possibile che anche il problema della *mano aliena* sia linguistico: la mano sinistra è controllata dall'emisfero destro, ma non potendo collegarsi ai centri del linguaggio il soggetto non riesce a esprimere linguisticamente ciò che pur sa di star facendo, né riesce con quella mano a eseguire ordini espressi tramite il linguaggio. Forse è così, forse non è così: l'unico modo per saperlo è capire prima l'esperienza soggettiva dell'avere una *mano aliena*: che si percepiscano i suoi movimenti come del tutto separati da sé, o sia invece una sensazione complessa e ambigua, una specie di *super-tic*, è rilevante, ma il filosofo dovrebbe vivere queste cose in prima persona per giungere a conclusioni fondate.

La libertà è presupposta dalla pratica scientifica: Einstein, Bell e Kochen

Incurante del pericolo, prendo un grammo di uranio purissimo. Tutti i suoi atomi, per tutti i parametri che posso analizzare, sono identici. Eppure migliaia di volte al secondo uno di questi atomi decade, espelle dal suo nucleo due protoni e due neutroni per diventare un atomo di torio. Se questi atomi sono strutturalmente identici, e il decadimento è in un qualche modo determinato dalla loro struttura, perché alcuni decadono e altri no, invece di decadere tutti insieme? Posso analizzare statisticamente il decadimento, posso individuare il *tempo di dimezzamento*, dopo quanto tempo metà degli atomi di una sostanza sarà decaduta, ma non posso sapere quando uno specifico atomo decadrà, pare un evento genuinamente casuale.

La casualità causa ansia tanto ai fisici quanto ai filosofi. Einstein stesso credeva che la meccanica quantistica fosse incompleta, che un giorno sarebbe stata ridotta ad una teoria deterministica. Da questa speranza nasce l'ipotesi delle *variabili nascoste*: il comportamento casuale di certi fenomeni è di tipo epistemologico, non ontologico, è dovuto al mio non conoscere i parametri che causano quel comportamento, ma se li conoscessi, se avessi accesso a queste variabili nascoste, potrei produrre una teoria deterministica.

C'è un esperimento mentale che sostiene l'esistenza delle variabili nascoste, ma prima di esporlo devo richiamare alla memoria due fenomeni della fisica quantistica, il *problema dell'osservatore* e l'*entanglement*.

Sono nella mia cameretta, al buio, so che di fianco all'armadio c'è una pianta ma non posso toccarla. Come posso sapere com'è fatta? Posso illuminarla con la torcia: devo in qualche modo *interagire* con la pianta, devo *immettere energia* nel suo sistema; se non posso interagire, non posso averne conoscenza. Al che mi chiedo: com'è la pianta quando *non* interagisco con essa? La pianta al buio potrebbe essere esattamente com'è quando la illumino: ma se questo è vero per gli oggetti macroscopici, per le particelle più piccole non è così. Io posso osservare una particella solo se interagisco con essa (tramite rilevatori, macchinari, i miei occhi o quel che si vuole), ma tale particella è così sensibile che qualsiasi interazione la modifica, le dà proprietà diverse da quelle che aveva prima dell'osservazione. È come se quella che mi appare come una pianta sotto la luce, al buio diventasse un dragone, e io non posso saperlo. Anzi, gli antirealisti affermano direttamente che la pianta, cioè la particella, non ha alcuna proprietà precisa prima dell'osservazione, la posizione della particella è *indefinita* finché non interagisco con essa: esiste una realtà indipendente dall'osservatore, o un albero che cade quando nessuno lo sente non fa alcun rumore? Questo è il *problema dell'osservatore*.

L'*entanglement* è un altro fenomeno inaspettato. Ho in tasca due particelle in stato di *entanglement*, e ora le separo di migliaia di anni luce, una la tengo qua e l'altra la spedisco in vacanza in un'altra galassia. Ebbene, se io osservo una particella, e così facendo cambio le sue proprietà, anche l'altra particella, non importa quanto lontana, cambia *istantaneamente*. Per la teoria della relatività nulla viaggia *nello spazio* più velocemente della luce, e quindi non è neanche possibile *trasmettere informazione* più velocemente della luce, gli effetti di un evento non possono propagarsi subito in tutto l'universo; eppure è come se dell'informazione passasse da una particella all'altra a velocità infinita, come se la particella lontana venisse a sapere all'istante che qualcosa è cambiato.

Einstein, assieme ai fisici Podolsky e Rosen, sviluppa un esperimento mentale, detto *paradosso EPR*, a favore della teoria delle variabili nascoste, poi rielaborato dal fisico Bohm; la versione che espongo ora è appunto quella bohiana.

Le particelle hanno una proprietà chiamata *spin*, difficile da definire (è come se girassero su sé stesse, ma in realtà non lo fanno). Prendo due particelle, *A* e *B*, e formo un *singoletto*, cioè una coppia in cui una particella ha lo spin di valore opposto a quello dell'altra. Divido quindi le particelle, ne spedisco una in un'altra galassia con posta prioritaria. Io posso non conoscere lo spin delle particelle, ma se so che formano un singoletto, allora misurando lo spin di *A* conosco automaticamente anche quello di *B* pur non avendolo mai misurato direttamente. Questo può essere ripetuto per un gran numero di proprietà in particelle *entangled*, anche quelle che potrebbero essere indefinite prima dell'osservazione.

Che ciò sia causato da un passaggio d'informazione a velocità superiore a quello della luce è inaccettabile per la relatività: ciò che determina il risultato della mia misurazione dello spin deve essere qualcosa di *vicino* o *interno* alla particella, qualcosa che ha il tempo di raggiungerla, e non un evento in una galassia lontana lontana: questo è il *principio di località*.

Come spiegare, allora, la mia possibilità di conoscere lo spin di *B* tramite un'osservazione su *A*? Devo sostenere il *principio di realismo*: la proprietà misurata è reale, indipendente dall'osservatore, e avrebbe un certo valore anche se nessuno interagisse con la particella. Cioè: poiché io conosco il valore dello spin di *B* anche senza osservarlo, esso non è creato dall'osservazione, bensì esiste indipendentemente; e se questo vale per *B* deve valere anche per *A*, e se vale per lo spin vale anche per le altre proprietà così misurate. Se i principi di località e realismo sono veri, allora la teoria quantistica, che li nega, deve essere incompleta, cioè devono esservi delle variabili nascoste e *locali* che spiegano i fenomeni osservati.

Questo esperimento mentale ne richiama un altro, quello del *gatto di Schrödinger*. Prendo una scatola contenente un contatore Geiger sensibilissimo, e vi immetto un atomo di uranio. Dopo un certo lasso di tempo, l'atomo potrebbe decadere come no, a caso: se rimane stabile, non succede niente, ma se decade, il contatore Geiger lo rileva e aziona un martello che rompe una fiala contenente veleno. Metto nella scatola un gatto: passato il tempo stabilito, se l'atomo sarà decaduto, troverò il gatto morto avvelenato, ma se non sarà decaduto, recupererò il gatto vivo. Il problema è che nell'*interpretazione di Copenhagen* della meccanica quantistica un atomo può essere descritto come esistente in una *sovrapposizione di stati*: per esempio, l'atomo può essere eccitato e non-eccitato assieme, anche se all'osservazione mostrerà solo uno dei suoi stati. Al che, se l'atomo può essere contemporaneamente decaduto e non-decaduto, anche il gatto deve essere morto e non-morto insieme finché non apro la scatola e faccio un'osservazione. Ma questo è assurdo: come ente macroscopico cosciente, il gatto o è vivo, o è morto. Quindi la meccanica quantistica è incompleta, è incompleta perché non assegna agli oggetti le proprietà che devono avere in un certo istante.

Il fisico Bell risponde a queste osservazioni con un teorema che è tra i più profondi di tutta la scienza.

Immagino ancora di avere un singoletto di particelle. Io posso misurare il loro spin *in più direzioni*: visualizzando la particella come su un piano cartesiano, io magari scelgo di misurare il suo spin lungo l'asse x dal basso in alto, mentre un fisico dell'altra galassia decide di misurarlo lungo l'asse y da destra a sinistra. Lo spin di una particella del singoletto è sempre opposto a quello dell'altra *se la misurazione è effettuata lungo lo stesso asse e nella stessa direzione*. Immagino ora che due fisici misurino accordandosi su asse e direzione, e facciano quattro tentativi:

	Tentativo 1	Tentativo 2	Tentativo 3	Tentativo 4
Particella A	↑	↓	↑	↓
Particella B	↓	↑	↓	↑

Il simbolo ↑ indica che si è osservato uno *spin-up*, ↓ uno *spin-down*, i due possibili *stati* dello spin (è come se le particelle ruotassero in direzioni opposte, non è davvero così ma per gli scopi della filosofia è inutile precisare).

Devo quindi calcolare la *correlazione*. In fisica è più semplice che in statistica: basta moltiplicare i valori di A e B (che saranno, per convenzione, + 1 per lo *spin-up* e - 1 per lo *spin-down*) tentativo per tentativo, e poi fare la media. In questo caso, ogni tentativo ha valore $1 \times - 1 = - 1$, quindi la media dei tentativi, cioè la correlazione, è sempre - 1.

Cosa succede, però, se i due fisici non si accordano su asse e direzione di misurazione? Se un fisico misura lo spin lungo l'asse x e l'altro lungo l'asse y , cioè due assi che formano un angolo a 90° , vi sarà il 50% di probabilità di ottenere lo stesso valore, e il 50% di ottenere valori opposti; per esempio:

	Tentativo 1	Tentativo 2	Tentativo 3	Tentativo 4
Particella A	↑	↓	↓	↑
Particella B	↑	↑	↓	↓

In tal caso, i tentativi 1 e 3 hanno valore $+1$, e i tentativi 2 e 4 -1 , quindi la correlazione sarà 0 .

Se invece i fisici misurano lungo lo stesso asse ma in direzioni opposte, otterranno sempre lo stesso risultato:

	Tentativo 1	Tentativo 2	Tentativo 3	Tentativo 4
Particella A	↑	↑	↓	↓
Particella B	↑	↑	↓	↓

In tal caso il valore di ogni tentativo è sempre $1 \times 1 = -1 \times -1 = 1$, e quindi anche la correlazione è 1 .

Se con un angolo a 0° (stesso asse, stessa direzione) la correlazione è -1 , e con angolo a 90° è 0 , viene naturale immaginare che con un angolo a 45° la correlazione debba essere $-0,5$, cioè che si debbano ottenere identiche misurazioni nel 25% dei casi. In effetti, tutte le teorie a variabili nascoste locali prevedono una correlazione di $-0,5$ o maggiore. La meccanica quantistica, invece, prevede una correlazione di circa $-0,7071$ (per la precisione $-\frac{1}{\sqrt{2}}$, cioè il coseno di quell'angolo), molto meno del minimo accettabile dalle altre teorie. Queste aspettative possono essere sottoposte a test empirico, com'è stato fatto più volte, con esperimenti sempre più raffinati e precisi. La conclusione è che le teorie quantistiche fanno previsioni più precise di quanto qualsiasi teoria con variabili nascoste locali possa fare.

In effetti, lo stesso Einstein non era convinto della versione di Podolsky, né ha mai sostenuto teorie a variabili nascoste locali se non per un breve periodo.

Il determinista ha ora tre modi per difendere la sua posizione. Il primo sta nel mantenere il principio di realismo ma nel rinunciare a quello di località, cioè può passare alla teoria delle variabili nascoste *non-locali*, proprietà che legano oggetti anche enormemente distanti e che possono quindi manifestare cambiamenti istantanei nelle due particelle. Il secondo sta, al contrario, nel mantenere la località rinunciando al realismo, può accettare che davvero le particelle non hanno certe proprietà prima dell'osservazione. Il terzo, l'unico che permette di mantenere entrambi i principi, è passare alla teoria della *superdeterminazione*: ogni teoria vista finora, anche quando deterministica, presuppone la *libertà di scelta*, presuppone che lo scienziato sia libero di scegliere come condurre i suoi esperimenti; in una teoria superdeterministica, invece, non c'è libertà ma esiste una correlazione tra il sistema misurato (la particella, l'universo), la scelta del metodo di misurazione (lungo quale asse misurare lo spin), e il risultato ($+1$ o -1).

Se c'è libertà di scelta, e i valori delle proprietà della particella A dipendono da come liberamente scelgo di effettuare l'osservazione, e i valori della particella B dipendono da quelli di A , allora i valori di B sono indeterminati finché non osservo A , e di conseguenza pare esserci dell'informazione che passa istantaneamente da A a B . In una teoria superdeterministica non è così: io sono determinato sin dall'inizio dell'universo a compiere l'osservazione in un certo modo, quindi a ottenere un certo valore di A , di modo che anche il valore di B resta determinato; di conseguenza, non c'è alcuna informazione che viaggia a velocità infinita, tutto è giustificato dalla determinatezza. Il problema ora è: come verificare la verità di località e realismo all'interno del superdeterminismo, e quindi la verità del superdeterminismo stesso?

Se io metto dell'idrogeno in un'ampolla, e osservo che la sua temperatura aumenta nel tempo, questo può essere dovuto a più *variabili*: potrebbe essere una proprietà intrinseca dell'idrogeno, potrebbe essere dovuto al fatto che l'ampolla era più calda del gas immesso, potrebbe dipendere da raggi cosmici che aumentano il livello d'energia delle particelle, o da movimenti dell'ampolla che a sua volta trasmette il moto agli atomi all'interno... Io però posso escludere man mano queste variabili: posso mettere uno schermo protettivo attorno all'ampolla, posso fermarla bene e raffreddarne l'esterno, attaccarle rilevatori di movimento e termometri, per poi vedere se la temperatura continua ad aumentare.

Questo non lo posso fare nel caso del superdeterminismo: qualsiasi risultato io ottenga dai test, posso giustificarlo con l'esistenza di variabili nascoste, determinate sin dall'inizio dell'universo, che determinano a loro volta le mie scelte sul come condurre gli esperimenti e il loro risultato. È un po' come se, posta l'ipotesi «esistono oggetti che cadono anche verso l'alto», all'obiezione «ogni oggetto osservato finora cade solo verso il basso» rispondessi superdeterministicamente «è così perché sono determinato a osservarli solo quando non cadono verso l'alto»: l'avversario non può più fare niente per dimostrare che mi sbaglio, rifiuterò qualsiasi osservazione contraria all'ipotesi dicendo che sono determinato a osservare quanto osservo *anche se* tali oggetti esistono. Ma per la scienza è inaccettabile spiegare i risultati dei test tramite il mio essere manipolato a osservare quei risultati: si tratta di una teoria infalsificabile, una teoria che non può essere dimostrata falsa neanche se fosse tale, e quindi non scientifica.

La comunità scientifica tende a rifiutare in modo compatto il superdeterminismo e a riconoscere che, se il superdeterminismo è falso, è improbabile (anche se non impossibile) che una teoria con variabili nascoste non-locali, o una teoria locale antirealista, possa offrire previsioni migliori di quelle della meccanica quantistica. Pare, insomma, che l'argomentazione di Einstein sia da rifiutare, e che davvero la fisica quantistica sia, in tal senso, completa.

Partendo dunque dal presupposto della libertà di scelta posso implicare qualche altra nozione rilevante per il libero arbitrio.

Bell, in contemporanea coi matematici Kochen e Specker, aveva già dimostrato un secondo teorema: aveva dimostrato che la precisione delle previsioni della meccanica quantistica è irraggiungibile non solo per qualsiasi teoria con variabili nascoste locali, ma anche per qualsiasi teoria con variabili nascoste *non-contestuali*.

Il *principio di contestualità* afferma che, quando misuro la proprietà di una particella, tale proprietà è associata non solo alla particella, ma anche a tutto l'apparato usato per la misurazione (che sia il mio occhio, macchinari vari o altro). È un po' come se le proprietà che io osservo nella pianta in camera (essere verde, fare la fotosintesi, avere le radici) dipendessero sia dalla pianta sia dal tipo di torcia usata per illuminarla. Come già intuito da Einstein, è possibile dimostrare che dal principio di contestualità si implica la non-località.

Il realismo richiede il principio di contestualità, realismo e non-contestualità sono in contraddizione. Infatti una teoria delle variabili nascoste, se è realista (cioè se afferma che il valore delle proprietà è sempre definito indipendentemente dall'osservazione), può cercare di rendere merito dei fenomeni quantistici tramite il principio di contestualità, cioè accettando che l'osservazione non rivela le proprietà della particella così com'erano l'istante prima ma bensì le modifica. Al contrario, se rifiuta il principio di contestualità, potrà sperare di fare previsioni buone quanto la teoria quantistica solo rifiutando anche il realismo, il che è indesiderabile. Oggi non-località e contestualità sono principi accettati da quasi tutti i fisici.

Dal principio di contestualità posso ora sviluppare questo ragionamento:

- *libertà di scelta*: posso condurre l'esperimento come voglio;
- *principio di contestualità*: l'atto di osservazione influenza le proprietà della particella;
- diversi modi di condurre l'esperimento influenzano in diverso modo le proprietà della particella;
- quindi, se il mio agire non è determinato dagli stati passati, neanche le proprietà *osservate* della particella lo sono, cioè è impossibile prevedere il risultato di un esperimento futuro *solo* sulla base di un qualsiasi stato precedente quell'esperimento (come la storia dell'universo, gli esperimenti passati, o altro).

Kochen, accompagnato dal matematico Conway, giunge a queste stesse conclusioni in maniera molto più rigorosa, tramite quello che ha scelto di chiamare direttamente *teorema del libero arbitrio*.

Ho due particelle, A e B . Posso svolgere un'operazione chiamata *misurazione del quadrato della componente di spin*: tutto ciò che devo sapere d'essa è che il suo risultato è 0 o 1. Come nell'esempio di Bell, posso effettuare la misurazione in più direzioni. La fisica mi dice che se misuro in tre diverse direzioni tra loro perpendicolari otterrò due volte 1 e una volta 0, in un qualche ordine: misurando nelle direzioni x, y, z , potrei ottenere «1, 1, 0» oppure «1, 0, 1» oppure «0, 1, 1». Le particelle A e B sono *gemellate*, cioè legate da un tipo di *entanglement* per il quale A e B , alla misurazione, daranno gli stessi risultati per le stesse direzioni: se misurando A nella direzione x ottengo 1, B pure mi darà 1 in quella direzione.

Secondo il determinista, il risultato della misurazione di A è determinato dall'*informazione* α (che include tutto ciò che può raggiungere la particella e potenzialmente influenzarla, dall'inizio dell'universo fino all'istante precedente la misurazione), e allo stesso modo quello di B lo è da β . Dirò, per esempio, che esiste una funzione $f(x; \alpha)$ che collega la misurazione nella direzione scelta x , dato α , ad un valore v (dove v è 1 o 0). L'esistenza di tale funzione costituisce la tesi determinista.

Si dà per scontato che in questo universo non vi sia retrocasualità: quindi tutto ciò che è compreso in α , precedendo la scelta della direzione da misurare, non può essere causato dalla scelta di x ; ed essendovi *libertà di scelta*, non può neanche causarla, quindi c'è totale indipendenza tra α e x (lo stesso valga per qualsiasi coppia di informazioni $\alpha, \beta...$ e scelte di direzione $x, y, z...$ proprie o altrui che siano).

Ora, c'è una configurazione di direzioni, detta *configurazione di Peres*, composta da 33 direzioni possibili lungo le quali effettuare la misurazione. Kochen aveva già dimostrato con Specker che è impossibile assegnare alle 33 direzioni un valore di 1 o 0 in modo tale che:

- ogni tripletta di direzioni x, y, z tra loro perpendicolari dia come risultati due 1 e uno 0 in un qualche ordine;
- due direzioni perpendicolari non diano mai entrambe 0;
- due direzioni opposte diano sempre lo stesso risultato.

Cioè, anche se la misurazione dà risultati sempre coerenti con le tre regole esposte, è impossibile che tutte le direzioni abbiano intrinsecamente un valore di 1 o 0 poi scoperto dalla misurazione, in quanto non posso assegnare tali valori fissi a tutte le direzioni rimanendo coerente con tutte e tre le regole; quindi, in tale configurazione, il valore del quadrato della componente di spin rilevato dalla misurazione non può essere indipendente, non può precedere la misurazione stessa. In altre

parole: la *configurazione di Peres* richiede il principio di contestualità, i valori devono variare con il contesto.

Dunque, io posso scegliere tre di queste 33 direzioni, purché perpendicolari, lungo le quali effettuare la misurazione su A , e un fisico di un'altra galassia sceglierà una direzione qualsiasi da misurare su B . Secondo la tesi determinista, il risultato della mia misurazione è determinato da $f(x,y,z;\alpha)$: dico cioè che il risultato ottenuto dipende dalle tre direzioni scelte e dall'informazione α . Per l'altro fisico la funzione invece è $f(x;\beta)$: stiamo quindi misurando una stessa direzione (x) ma in due luoghi e in due momenti diversi (io rispondo ad α , lui a β). Nei due casi, il risultato *deve* essere lo stesso.

Affinché ciò sia vero, in primo luogo, il risultato della misurazione di x deve essere indipendente da quali altre perpendicolari y o z posso aver scelto: infatti, qualsiasi siano le mie scelte, queste non possono influenzare il risultato della misurazione dell'altro fisico, le mie scelte e i risultati delle mie misurazioni sono al di fuori di β (anche perché è troppo lontano, gli effetti delle mie azioni non fanno in tempo a raggiungerlo), e poiché dobbiamo avere gli stessi risultati, come y e z non possono influenzare il suo risultato per x , così non possono influenzare il mio. In secondo luogo il risultato deve essere indipendente anche dall'informazione in α e β , perché dobbiamo ottenere lo stesso risultato qualsiasi sia l'informazione passata. Di conseguenza, è irrilevante pure chi dei due compia la misurazione per primo: se io misuro dopo l'altro fisico, devo ottenere il suo stesso risultato, ma se misuro prima di lui, comunque devo ottenere lo stesso risultato che *avrebbe* ottenuto se *avesse* misurato prima di me.

In termini più formali: posto per ipotesi che $f(x;\beta)=1$, è necessario che $f(x,y,z;\alpha)=1$ per x qualsiasi siano i valori di y , z e α , dunque qualsiasi siano le direzioni scelte e qualsiasi siano le informazioni passate in grado di raggiungere A (non importa se la misurazione di B sia passata o futura rispetto ad A , posso modificare α per portare la misurazione avanti e indietro nel tempo, ma deve dare comunque 1). Quindi il risultato di una misurazione può dipendere solo dalla direzione scelta: $f(x,y,z;\alpha)=f(x)$, la funzione assegna un valore a qualsiasi direzione a prescindere da ogni altro fattore.

Questo vuol dire che il risultato di una misurazione non varia col contesto, viola il principio di contestualità. Ma ciò è dimostrato impossibile per la *configurazione di Peres* dalla quale sto scegliendo le direzioni. Quindi o non ho libertà di scelta e non posso davvero scegliere liberamente tra tutte le 33 direzioni, oppure il risultato della misurazione non è determinato da una qualche funzione. Cioè: se io sono libero di scegliere come misurare le proprietà di una particella, anche le proprietà che osserverò si comporteranno in modo libero. Se io sono libero, anche le particelle lo sono.

Questa argomentazione non vuole dimostrare il libero arbitrio umano: dice invece che il libero arbitrio è una possibile spiegazione delle caratteristiche della meccanica quantistica, e quindi che il libero arbitrio è abducibile dall'indeterminatezza osservata nelle particelle (le quali spiegano a loro volta l'origine del libero arbitrio umano: la mia libertà deriva dalla libertà delle mie particelle).

Ovviamente questo sarà insufficiente a convincere un determinista ad abbandonare la sua posizione: l'abduzione non è cogente, e la definizione di libero arbitrio adottata (cioè *non essere determinati dagli stati passati*) descrive qualcosa che non può garantire vera libertà. Se però questo determinista è anche un naturalista, il fatto che la comunità scientifica tenda a rifiutare il superdeterminismo (che è l'equivalente fisico del determinismo filosofico) rimane significativo: per rispettare i dati della fisica anche il filosofo dovrebbe mantenere l'assioma della libertà di scelta, a patto che altre teorie scientifiche altrettanto autorevoli non puntino in direzioni diverse.

La libertà è necessaria alla conoscenza

Ho scritto di Tolstoj, ora mi viene da pensare a Tolkien, l'autore del *Signore degli Anelli*. Tolkien faceva parte di un gruppo letterario, gli *Inklings*. Le loro riunioni avvenivano nella stanza dell'autore delle *Cronache di Narnia*, C. S. Lewis, un uomo la cui esistenza è il motivo per cui ho dovuto costantemente specificare il nome di David Kellogg Lewis. Tolkien e C. S. Lewis, consci del fatto che nessuno avrebbe scritto il genere di romanzo da loro desiderato, sono entrambi autori di epopee fantasy. Sono anche entrambi autori cristiani: C. S. Lewis non era credente, ma Tolkien lo ha convinto a convertirsi. Ed entrambi, oltre che per romanzi e racconti, sono noti anche per lavori più tecnici: Tolkien era filologo e linguista, C. S. Lewis invece è famoso in ambito protestante come teologo.

Per me, il nome di C. S. Lewis è legato anzitutto all'*argomentazione per ragione*, il suo più importante contributo teologico e filosofico. Ma non l'ha inventata lui, l'argomentazione si trova anche in molti altri autori: eccola in G. K. Chesterton scrittore e apologeta, in C. E. M. Joad pensatore e star radiofonica, in Richard Taylor filosofo e apicoltore... Pur nascendo in ambito religioso, l'argomentazione è stata difesa anche da non credenti: la sua essenza è già presente nel teorema di Bell, la si trova negli attacchi al superdeterminismo del fisico Nicolas Gisin, appare in Popper che ne dà una delle formulazioni più convincenti, ed è chiara anche nel biologo e statistico ateo J. B. S. Haldane che affermava «se i miei processi mentali sono determinati totalmente dal moto dei miei atomi nel mio cervello, non ho ragione di credere che le mie credenze siano vere, e quindi non ho ragione di credere che il mio cervello sia composto di atomi». In effetti, le versioni migliori di questa argomentazione sono laiche e neutrali; si possono quindi abbandonare i sospetti, perché l'argomentazione in sé non incarna interessi apologetici o altri *bias* religiosi.

Esporrò dunque più versioni; non tutte sono state intese dai loro autori come rilevanti per il problema del libero arbitrio, alcune avevano in origine altri scopi, ma una conclusione libertarista è facilmente implicabile.

C. S. Lewis intendeva la sua versione come una prova dell'esistenza di Dio, o meglio come una difesa della credibilità dei miracoli. Posso esprimerla così:

- gli esseri umani sono razionali;
- i processi naturali, come l'attività chimica dei neuroni o l'evoluzione, sono irrazionali;
- un processo irrazionale non può formare un pensiero razionale: se l'attività irrazionale dei neuroni è stata creata da un irrazionale processo evolutivo, allora non posso avere credenze razionali;
- quindi ciò che è razionale è irriducibile a processi irrazionali: la razionalità umana non può derivare da processi naturali.

Perché questo dovrebbe essere rilevante per il libero arbitrio? Il determinista pensa ai processi naturali del cervello come determinati, e da ciò inferisce che anche tutto il comportamento riducibile al cervello debba essere determinato; ma l'argomentazione mostra che tale conclusione è ingiustificata, perché se il comportamento umano fosse riducibile a processi naturali determinati non potrebbe esservi razionalità, e senza razionalità non si giustificano le credenze.

Come spesso accade per le prove dell'esistenza di Dio, anche questa fu accolta con sospetto dai pensatori cristiani (meno, invece, dagli atei). Le più forti obiezioni arrivano da Anscombe, che ne

offre tre. Prima: un processo razionale aumenta la probabilità di pervenire a credenze vere, un processo irrazionale la diminuisce, ma un processo naturale non fa né l'altra, i processi naturali sono non-razionali (e non irrazionali come dice C. S. Lewis). Seconda: l'argomentazione suggerisce che, se tutto il pensiero fosse spiegabile tramite cause naturali, allora sarebbe impossibile produrre un ragionamento *valido* (cioè un ragionamento in cui da premesse vere si implica correttamente una conclusione vera); questo vorrebbe dire che le cause naturali producono *sistematicamente* ragionamenti invalidi, ma non c'è motivo per il quale dovrebbe essere così, le cause naturali possono produrre sia ragionamenti invalidi che validi, almeno per caso. Terza: ciò che costituisce una *spiegazione valida* varia col contesto, dire che bevo *perché ho sete* è una buona spiegazione nella vita quotidiana ma non in biologia cellulare (che invece spiegherà il comportamento tramite processi biochimici); C. S. Lewis, però, pare accettare solo un tipo di spiegazione, considerando irrazionali tutte le altre, il che ha l'indesiderabile conseguenza di rifiutare un gran numero di scienze.

C. S. Lewis riconosce senza problemi la correttezza di queste obiezioni che, lungi dall'essere distruttive, sono superabili apportando piccole modifiche al ragionamento:

- gli esseri umani sono razionali;
- i processi naturali (non-liberi) sono *non-razionali*;
- un processo *non-razionale* non può formare un pensiero razionale in quanto non permette di produrre *in modo sistematico* ragionamenti validi (così il ragionamento umano rimane ingiustificato);
- quindi qualsiasi spiegazione naturale (non-libera) della credenza razionale (esempio: «lo credo perché i miei neuroni si comportano così») è invalida in quanto i processi naturali (non-liberi) *non aumentano la probabilità che la mia credenza sia vera*.

Un'obiezione interessante a questa versione è quella del computer. Oggi esistono algoritmi che, sfruttando la statistica, apprendono in autonomia: sottoposti a stimoli ripetuti, li analizzano, modificano man mano il loro stesso codice, e completato questo periodo di addestramento possono manifestare un comportamento complesso e all'apparenza intelligente, come guidare una macchina senza bisogno di assistenza umana. Un computer non è razionale, si limita ad eseguire l'algoritmo, il quale a sua volta è solo un circuito attraverso il quale passa elettricità; eppure i computer raggiungono performance elevate in campi che paiono richiedere intelligenza, i computer sono oggi molto più abili di qualsiasi umano in pressoché tutti i più complessi giochi da tavolo. Se un computer, funzionando tramite processi non-razionali, può agire razionalmente, perché non potrebbe essere lo stesso per l'uomo?

Il problema di questa obiezione sta nell'intenzionalità. Kripke, discutendo la teoria della traduzione di Quine, mostra che un algoritmo non può costituire il fatto di un significato. Se lo schermo di un computer mi mostra una parola, io non posso capire cosa il computer *intenda* analizzando il suo algoritmo, perché l'algoritmo stesso (qualsiasi esso sia) va interpretato; osservando i suoi circuiti, non posso capire se quello che mostrano è l'algoritmo che *dovrebbero eseguire*, o se invece il computer esegua l'algoritmo corretto solo quando i suoi circuiti sono bagnati. Putnam, da antirealista, trae le sue conclusioni e afferma che un computer non può intendere alcunché, non può *riferirsi* a nulla che sia al di fuori dei suoi circuiti, è come una formica che zampettando a caso finisce per tracciare sulla sabbia quella che *io* riconosco come una parola.

In Searle si trova un esperimento mentale che conferma la validità di queste conclusioni anche per il realista. Un uomo, che non conosce il cinese, viene chiuso in una stanza. Ci sono qui delle carte con disegnati ideogrammi cinesi e un manuale che spiega come usarle. Dall'esterno, qualcuno

immette un foglio con domanda in cinese. L'uomo, anche se non capisce gli ideogrammi, può ricercare le domande nel manuale e comporre una risposta da spedire all'esterno. Da fori, si vede solo la stanza che, ricevendo una certa domanda, risponde in modo adeguato. Mi chiedo: chi è che conosce il cinese, lì? Non certo l'uomo. Alcuni dicono *la stanza nel suo complesso* (uomo, manuale e ideogrammi insieme), ma è una tesi azzardata, la stanza non ha una mente, e quindi neanche conoscenze, solo l'uomo può averne. Quindi nessuno sa il cinese, la stanza *sembra* conoscerlo, ma in realtà le sue parole non vogliono riferirsi ad alcunché. Posso trasformare la stanza in un robot con braccia meccaniche comandate dall'uomo affinché tocchino gli oggetti di cui parla (dà così l'illusione del riferimento), ma non cambierebbe niente. Un computer è così: non ha intenzionalità, cioè non può riferirsi agli oggetti del mondo.

Quindi sì, un computer può manifestare comportamenti efficienti ed esprimere quelle che riconosco come verità, ma comunque non può essere razionale nel modo in cui lo è l'uomo perché, non avendo intenzionalità, non può neanche avere conoscenze. L'obiezione non funziona.

Bene, un primo scoglio è superato. Ma attenzione, si alza tra le onde una nuova critica. L'argomentazione continua a implicare che una credenza razionale non può avere una giustificazione irrazionale; per esempio, alla domanda «perché hai paura dell'altezza?» (o «perché *credi* che stare in alto sia pericoloso?») rispondere «l'evoluzione mi ha tramandato questa paura molto comune» non è certo razionale, lo è invece «perché so che se cado muoio», questo giustifica la mia paura. Posso obiettare però che, se ho buone ragioni per credere qualcosa, il modo in cui queste ragioni sono state prodotte è irrilevante: il sapere che se cado muoio riesce a giustificare la mia paura, ma tale giustificatore può ben essere prodotto dall'evoluzione o da un certo funzionamento determinato del mio cervello, e questo non distrugge affatto la razionalità della mia paura. Come salvare ora l'argomentazione?

Plantinga, potenziando l'argomentazione di C. S. Lewis grazie alla logica formale e alla statistica, sviluppa una sua versione volta a sostenere l'autocontraddittorietà del naturalismo:

- le *facoltà conoscitive* umane (i cinque sensi, il ragionamento...) sono *affidabili*, cioè sono un buon mezzo (anche se non infallibile) per produrre credenze vere;
- queste facoltà conoscitive sono prodotte dall'evoluzione naturale;
- l'evoluzione naturale è un processo senza scopo (non guidato), che premia le capacità di sopravvivenza e non la capacità di produrre credenze vere;
- quindi, se le facoltà conoscitive sono prodotte dall'evoluzione, la loro affidabilità deve essere bassa o inconoscibile (e, usando un *affidabilismo internalista*, se è inconoscibile, non sono giustificato a ritenerle affidabili, quindi non possono neanche giustificarmi e produrre conoscenza);
- ma le facoltà conoscitive umane sono affidabili, quindi o non sono il prodotto dell'evoluzione naturale, o l'evoluzione naturale non è un processo senza scopo che premia solo la sopravvivenza.

In questa versione affidabilista il modo in cui si produce il giustificatore diventa rilevante: per Plantinga avere delle buone ragioni è insufficiente, servono buone ragioni *affidabili*, le credenze prodotte da processi inaffidabili non possono più giustificarmi. E se non si concorda, si può comunque cambiare teoria epistemologica: tutti i modelli che accettano un quarto principio (oltre a credenza, verità, giustificazione), in particolare quelli esternalisti che richiedono la presenza o assenza di certi elementi del contesto, sono tali da rendere le semplici buone ragioni insufficienti alla conoscenza.

Plantinga è preso di mira da una pletera di nuove critiche, ma quasi tutte finiscono fuori bersaglio: molti lo accusano di aver dimostrato che l'evoluzione ha avuto un risultato inaspettato, e non che è falsa (ma infatti Plantinga non rifiuta l'evoluzione); altri dicono che, se vuole usare Dio per spiegare questo risultato improbabile, allora deve anche considerarlo causa di tutti gli altri eventi casuali come il decadimento radioattivo (e Plantinga, esagerando, è disposto a farlo, ma ciò non è affatto implicato dall'argomentazione).

Putnam offre un esperimento mentale che porta alle stesse conclusioni del ragionamento di Plantinga. Nel folto della foresta scopro un pulsante che, se premuto, fa comparire dal nulla una torta. Il portare una torta alla mia tribù è un'azione che aumenta le sue probabilità di sopravvivenza; io credo correttamente queste cose e desidero che la mia tribù sopravviva, così l'evoluzione naturale premierà il mio comportamento tramandando alle future generazioni le mie credenze e i miei desideri, che hanno come effetto la tendenza a premere questo benefico pulsante. Di contro, il mio amico sostenitore dell'estinzione volontaria vuole ridurre le probabilità di sopravvivenza della tribù, e crede erroneamente che portare torte sia un modo per realizzare questo scopo: la selezione naturale, però, tramanderà ugualmente le sue credenze e i suoi desideri, perché questi portano comunque a premere il pulsante e quindi ad aumentare le probabilità di sopravvivenza. Cioè: l'evoluzione seleziona i comportamenti e non le credenze in sé, quindi non può né favorire (se vera) né sfavorire (se falsa) una credenza a meno che non abbia effetti comportamentali; ma poiché uno stesso comportamento può essere prodotto sia da credenze vere che da credenze false (e sempre con l'influenza di preferenze, tratti di personalità e altri fattori tutti indipendenti dalla verità della credenza), i processi creati dall'evoluzione non paiono affidabili.

Epperò Putnam rifiuta subito queste conclusioni. Da antirealista ed esternalista, appoggia una nozione particolare di conoscenza derivata dalla sua *teoria causale del riferimento*. Poiché il significato delle parole ha un legame causale con l'oggetto a cui si riferiscono, e le credenze sono attitudini proposizionali (quindi dipendenti dal linguaggio), anche le credenze sorgono tramite l'interazione con la realtà. È ovvio che, tra tutte le credenze false, una grandissima maggioranza porta a comportamenti deleteri per la sopravvivenza, e solo una piccola fetta porta comunque a comportamenti efficaci. Di conseguenza, l'idea che l'interazione con l'ambiente produca una fetta rilevante di credenze false che, per puro caso, favoriscono la sopravvivenza, è altamente improbabile. Se l'evoluzione non producesse una maggioranza di credenze vere, be', la sopravvivenza umana sarebbe un miracolo inspiegabile.

Con ciò, Putnam non voleva certo rifiutare l'argomentazione di Plantinga, visto che scriveva una dozzina d'anni prima di lui; eppure questa è la critica più centrata all'argomentazione per ragione. Un'altra formulazione della stessa critica la presenta il filosofo della mente e del linguaggio Jerry Fodor: secondo lui, la teoria dell'evoluzione offre una *storia* su come possano essere andate le cose, e questa *storia* spiega chiaramente perché il meccanismo dell'evoluzione *può* produrre una maggioranza di credenze vere, e quindi perché le facoltà conoscitive umane possono essere affidabili; certo poteva anche produrre facoltà inaffidabili, ma voler credere così sarebbe solo un contesto scettico, un dubbio iperbolico. Queste obiezioni meritano attenzione, e valgono anche per la versione di C. S. Lewis.

Il problema centrale dell'argomentazione così com'è stata esposta finora è il suo presentarsi come un attacco al naturalismo. Il filosofo David Kyle Johnson, commentando una ricerca di Reppert, nota che l'argomentazione per ragione parte dal presupposto che il naturalismo neghi l'esistenza della mente, e questo è un errore perché molti naturalisti rifiutano la riduzione della mente alla materia. Similmente, il filosofo Angus Menuge difende l'argomentazione partendo dal presupposto che il naturalismo sia incompatibile con il libero arbitrio; ma, come lui stesso nota, i naturalisti

rispondono proponendo il compatibilismo, e chi crede nel compatibilismo crede nel libero arbitrio (in quanto compatibile col determinismo): per difendere l'argomentazione allora si dovrebbe dire che il naturalismo è incompatibile col compatibilismo, e questo sarebbe falso. Per superare queste obiezioni, devo ora esporre una versione personale dell'argomentazione per ragione, una versione che si concentri sulla compatibilità tra determinismo e giustificazione della credenza ma senza badare al naturalismo:

- se il determinismo è vero, allora, sin dalla nascita dell'universo, esiste un unico futuro possibile;
- quindi ciò che io crederò in futuro è già determinato sin dalla nascita dell'universo: io crederò le proposizioni che sono determinato a credere qualsiasi sia il loro valore di verità;
- ma una credenza è giustificata solo se indefettata (solo se il contesto è scevro da elementi che, fossero conosciuti, mi farebbero riconoscere che la mia giustificazione è invalida);
- se sono in un contesto in cui sono determinato a credere vero ciò che credo indipendentemente dal suo valore di verità, allora le mie credenze non possono essere indefettate, cioè in tale contesto la giustificazione è impossibile;
- se l'universo contenesse processi casuali, crederci ciò che credo *a caso*, e questo ugualmente rende impossibile la credenza indefettata;
- quindi, un universo costituito solo da processi deterministici e/o casuali costituisce un contesto scettico nel quale la conoscenza è impossibile, e va rifiutato come tutti gli altri contesti scettici.

Analizzo. Un contesto scettico è tale che, se si dà, non posso mai essere giustificato, ma se non si dà, io non posso saperlo. Se il mondo fosse determinato, ognuno potrebbe essere determinato a credere che non lo sia; posso offrire delle prove a favore della non-determinatezza, ma potrei ritenere queste prove *prove* solo perché sono determinato a ritenerle tali, magari in realtà non possono provare alcunché. Se c'è un unico futuro possibile, e quindi credo tutto ciò che sono determinato a credere senza poter credere altro, la mia credenza risulta sempre indipendente dalla verità: io credo ciò che credo «perché così è deciso sin dall'inizio dell'universo», ma questo non è un giustificatore valido. E per questo, se io sapessi di essere determinato, riconoscerei invalida ogni mia giustificazione.

Allora: come in sogno posso esperire tutto ciò che esperisco da sveglio senza esserlo, così potrei essere determinato ad esperire tutto ciò che esperirei se fossi libero pur non essendolo; come in sogno crederci che qui ci sia un tavolo anche se in realtà non ci fosse, così se fossi determinato crederci ciò che credo anche fosse tutto falso. Perciò, fossi determinato, non potrei saperlo, e sarebbe impossibile avere una credenza indefettata (e quindi conoscenza).

Ma se il determinista credesse davvero di essere determinato a credere cose false, sarebbe paradossale, il determinista *deve* credere di essere determinato a credere una maggioranza di cose vere. Infatti ciò che Putnam e Fodor obietterebbero è che, anche se sono determinato, grazie all'evoluzione lo sono in modo tale da dover avere credenze vere; e se è così, non crederci vero qualcosa anche fosse falso, non sono un orologio rotto che segna una certa ora anche quando sbagliata, il meccanismo dell'evoluzione mi guida affinché io segua il vero, quindi il determinismo è insufficiente a invalidare la giustificazione.

Questa visione è però inaccettabile dal principio di indefettità. Io parto da due principi:

1. *internalismo*: la mia giustificazione è valida se ho accesso al giustificatore;

2. *esternalismo*: affinché la mia giustificazione sia invalida è sufficiente che esistano elementi del contesto che la annullano.

Analizzo quindi tre esempi:

1. ho un orologio rotto, ma Dio mi guida in modo tale che io lo guardi sempre e solo quando segna l'ora corretta;
2. sto sognando, ma i miei sogni sono guidati da Dio affinché mi mostrino solo cose vere;
3. sono determinato, ma ho un chip nel cervello programmato in modo tale da farmi credere solo a verità di fatto.

Il determinista fodoriano, partendo dal primo, può ragionare così:

- P1** Credo siano le due perché l'orologio funziona in modo tale da indicare l'orario corretto;
- P2** ma l'orologio è rotto;
- P3** se l'orologio è rotto, allora, se l'ora indicata fosse falsa, io la crederei comunque (per questo può annullare la mia giustificazione);
- P4** ma Dio mi determina in modo tale che io guardi l'orologio solo quando segna l'ora corretta;
- C** quindi l'ora che io vedo non è falsa, quindi sono giustificato in quanto il fatto che l'orologio sia rotto non è un elemento sufficiente ad annullare la giustificazione.

Ma il ragionamento non è valido. L'orologio è il mio giustificatore perché funziona in un certo modo, e questo funzionamento è il suo metagiustificatore. Cosa annullerebbe questa metagiustificazione? Il fatto che l'orologio non funziona; il fatto che «non funziona e lo guarerò anche quando l'ora è falsa» pure lo annullerebbe, ma non è necessario.

Infatti, immagino che, in tutta la mia vita, io abbia guardato un orologio rotto solo una volta, e quell'unica volta l'orologio segnava l'ora corretta. Dal fatto che io non abbia mai guardato un orologio che segna l'ora sbagliata posso dedurre che, in quel contesto, ero giustificato anche se ho guardato un orologio rotto? Ovviamente no: per essere giustificato, avrei dovuto sapere che non ho mai guardato un orologio che segna l'ora scorretta, ma non lo sapevo.

In altre parole: se io conoscessi tutto il contesto, saprei sia che l'orologio è rotto (e quindi non sono metagiustificato in virtù del suo meccanismo) sia che esista un Dio che mi guida (e quindi che sono metagiustificato in virtù di questo fatto); se però io non ho accesso al fatto che Dio mi guida non ho questa nuova metagiustificazione, mentre il fatto che l'orologio sia rotto è sufficiente ad annullare la metagiustificazione corrente. Lo stesso vale per gli altri esempi: affinché i miei sogni siano mezzi di conoscenza è insufficiente che Dio li guidi, è necessario che io *sappia* che li guida; se io *sapessi* di essere determinato da un tale chip impiantato nel mio cervello, potrei essere giustificato, ma se non lo so non posso avere conoscenza.

Come in sogno potrei vedere per lo più cose false, così potrei essere determinato a credere falsità. Dunque, se vendendo un tavolo so che lì c'è un tavolo, o non sto sognando, o conosco un meccanismo che rende le percezioni affidabili anche in sogno; o non sono determinato, o so che sono determinato a credere come giustificatori solo mezzi affidabili. Ma in sogno non posso essere giustificato a credere in questo meccanismo; e se sono determinato, non posso essere giustificato a

credere di essere determinato a credere solo cose vere. Quindi, se sono determinato a credere solo cose vere, comunque non ho conoscenza. Ecco allora l'ultima versione dell'argomentazione:

- P1** sono giustificato solo se ho accesso ai giustificatori e non ci sono elementi del contesto che, fossero conosciuti, li annullerebbero;
- P2** per ipotesi, io sono determinato a credere cose vere;
- P3** se sono determinato a credere cose vere, allora sono determinato;
- P4** l'essere determinato è sufficiente ad annullare la giustificazione (non è necessario che sia determinato e lo sia a credere anche cose false);
- C1** quindi, se credo x , e sono determinato, non sono giustificato a crederla;
- P5** se io sapessi che non sono determinato a credere una x falsa, allora sarei giustificato;
- C2** ma, se sono determinato, non posso sapere che non sono determinato a credere una x falsa, e perciò non posso essere giustificato.

In virtù di questa ultima versione l'obiezione di Putnam e Fodor viene scartata: l'evoluzione mi spiega come sia possibile che io abbia una maggioranza di credenze vere anche fossi determinato, ma se sono determinato non posso avere giustificatori della mia credenza nell'evoluzione. Proprio perché l'evoluzione è vera, e davvero è avvenuta in modo tale da garantire la capacità di conoscere, e io sono certo di essere giustificato a credere in essa, non posso essere determinato.

Di recente lo psicologo Lockie (2018) ha difeso una *argomentazione trascendentale*, d'ispirazione epicurea, molto simile a quella da me proposta. Zanetti (2019), ricercatore dell'università di Bologna, ha sollevato un'obiezione notando che la teoria di Lockie presuppone un'epistemologia internalista e deontologica per la quale la giustificazione sta nell'assolvimento degli obblighi cognitivi: come le ragioni etiche mi dicono che *devo* fare qualcosa, così le ragioni epistemologiche mi dicono che *devo* ragionare in un certo modo. Anche se sono disposto a riconoscere l'esistenza di uno stretto legame tra etica e conoscenza, la mia versione dell'argomentazione non necessita di un'epistemologia di questo tipo.

Detto tra parentesi: i determinismi non-nomologici sono compatibili con la libertà

Finora ho parlato solo del *determinismo nomologico*, quel determinismo associato alle scienze naturali e caratterizzato dall'esistenza di un unico futuro possibile. Secondo tale determinismo, dato un universo con certe leggi fisiche e una certa distribuzione degli enti, da un qualsiasi suo istante è possibile calcolare ogni altro istante passato e futuro; quindi qualsiasi altro universo uguale a questo in un suo istante, lo sarà anche in ogni altro istante. L'argomentazione per ragione potrebbe rifiutare il determinismo nomologico; ci sono però molti altri tipi di determinismo che possono risultare problematici dal punto di vista epistemologico.

Penso anzitutto al *determinismo linguistico*. Come diceva Wittgenstein, «i limiti del mio linguaggio significano i limiti del mio mondo»: per questo uno dei miei scrittori, James Joyce, appena ventenne parlava già più lingue dei suoi anni, perché voleva fuggire dalla *gabbia culturale* nella quale era nato. Ma perché il linguaggio è una gabbia?

Secondo vari filosofi analitici, non c'è pensiero senza linguaggio: il linguaggio non è uno strumento col quale esprimere il pensiero, è il pensiero che è un'operazione sulle proposizioni espresse dal

linguaggio; di conseguenza, io posso comprendere ciò che gli altri tentano di esprimere solo tramite le mie categorie, solo tramite il mio linguaggio. Riprendendo l'esempio fatto in etica, le differenze tra «bene» e «buono», tra «un buon uomo» e «un uomo buono», sono difficilmente esprimibili in inglese: l'essere madrelingua italiano mi dona un accesso intuitivo a queste differenze; per un madrelingua inglese giungere alle mie stesse intuizioni potrebbe essere impossibile, e anche non lo fosse comunque gli sarebbe molto più difficile di quanto lo sia per me. In questo senso, il linguaggio determina e delimita le possibilità del mio pensiero, l'aver una certa lingua madre mi incapsula offrendomi certe categorie concettuali e non altre, mi predispone a pensare in un certo modo e ad avere quindi un certo ritratto del mondo.

Passo al *determinismo geografico*. Lo stupefacente Jared Diamond è un raro *biogeografo*: espertissimo di biologia, nei suoi libri incrocia storia, antropologia, psicologia, sociologia, economia e altre materie nel tentativo di comprendere l'origine dei comportamenti collettivi. Domanda: perché è stata l'Europa a scoprire l'America, e non l'America a scoprire l'Europa? Perché la civiltà europea ha conquistato il mondo, e le civiltà precolombiane sono scomparse? La risposta di Diamond è geografica: in Europa ci sono animali addomesticabili utili all'agricoltura e alla pastorizia (non così in America, un alpaca non vale una mucca); l'Europa si estende da Est a Ovest rimanendo all'interno di una certa fascia climatica, il che facilita viaggi e circolo d'informazione (difficili invece in America, che passa attraverso tutte le fasce climatiche); in Europa flora, fauna e clima ideali per l'agricoltura, permettendo la produzione di *surplus* di risorse da conservare o investire, favoriscono lo sviluppo di classi sociali agiate che, invece di preoccuparsi per la sopravvivenza, possono viaggiare, creare, inventare, ricercare, sviluppare tecnologia (la tecnologia precolombiana invece non era molto impressionante); l'alta densità abitativa europea facilita il diffondersi delle malattie, ma questo permette anche lo sviluppo un sistema immunitario forte (infatti le civiltà precolombiane sono state sterminate dalle epidemie); infine, tale densità abitativa, costringendo a lottare per una risorsa limitata qual è la terra, infonde nelle popolazioni uno spirito violento, guerriero, imperialista, che si traduce nello sviluppo di uno Stato centralizzato forte in grado di coordinare azioni militari collettive (le civiltà sviluppatasi in abbondanza di terra, invece, tendono alla disorganizzazione).

Altro esempio a me caro: perché c'è stato un impero romano, invece di un impero etrusco? La fortuna di Roma è stata l'aver accesso diretto alle saline delle foci del Tevere in un'epoca nella quale il sale era un bene indispensabile alla conservazione del cibo: proprio questa posizione strategica le ha permesso di svilupparsi e arricchirsi velocemente sin dai primi anni. Roma non era determinata a sfruttare questa risorsa, poteva non farlo, ma tutte le altre popolazioni erano invece determinate a non sfruttarla (non ne avevano accesso). Il capitare nel giusto territorio segna il destino: il territorio determina ciò che una civiltà può fare e può diventare.

Ma l'influenza del territorio va al di là di questo. Penso a Talete, il filosofo per il quale tutto discende dall'acqua. Talete è un greco, è sempre vissuto con mari, navi, isole davanti agli occhi; fosse nato in Tibet, avrebbe pensato le stesse cose? Probabilmente no. Una mitologia ricca di mostri marini e leggendari navigatori può caratterizzare le tribù di un arcipelago, ma difficilmente la troverò tra clan di pastori d'altura; le popolazioni dei climi freddi, stimolate dai rigori invernali, potrebbero sviluppare una maggior propensione alla collaborazione, ma solo tra le popolazioni delle fasce climatiche più calde potrò trovare nudità e assenza di abitazioni chiuse (ci sono tribù che vivono sotto lunghe tettoie, per esempio).

Questa è una visione, direi, *anti-continentale*: nella filosofia continentale la cultura influenza l'interpretazione umana del mondo naturale, ma nel determinismo geografico è il mondo naturale a influenzare la cultura e a determinare i suoi elementi. Io sono italiano, ma non perché l'Italia

esista in modo fattuale; lo sono perché, sin da quando sono nato, ho fatto esperienza di un certo ambiente, ho visto coi miei occhi certi monumenti, certi stili architettonici, certi animali, e queste esperienze mi accomunano alle persone che vivono vicino a me. Per questo, ancora più che italiano, sono veneto, e ancora più che veneto sono trevigiano, ho molte più esperienze in comune con gli abitanti di Meduna di Livenza che con quelli di Palermo o di Berlino. E queste esperienze, queste cose che ho visto attorno a me, hanno infine determinato i miei valori e il mio pensiero, rendendomi ciò che sono: fossi nato in Corea del Nord, non farei filosofia analitica e scriverei invece libri sul *Juche*.

L'equivalente continentale del determinismo geografico è il *determinismo tecnologico*, associato al marxismo. L'anno scorso, resomi conto di passare troppo tempo online, ho pensato di rendere il mio cellulare meno attraente impostando lo schermo in bianco e nero. Dopo pochi giorni, mi sono accorto che avevo iniziato a *pensare* in bianco e nero, e la visione di oggetti molto colorati mi stupiva come un'anomalia. Tanta è l'influenza di questo strumento sulla mia mente. Trattare l'attività online al pari della vita *vera* sembrava un'esagerazione fino a pochi anni fa, ora è normale e necessario, sono legalmente responsabile di ciò che scrivo su Facebook. Se, a causa di internet, la mia generazione ha uno *span d'attenzione* molto minore di quello della generazione di mio nonno, come poteva essere la mente di un uomo medievale? Forse ricordava molte più cose, era indifferente alle immagini, più attento ai suoni... e se è così, come sarà la mente dell'uomo del futuro, quando si impianterà chip nel cervello per espandere le sue capacità? La tecnologia davvero cambia la società e la mente, e così contribuisce a determinare la mia esperienza del mondo.

Devo parlare anche del *determinismo storico*. Ve ne sono due tipi, religioso e laico. Dal punto di vista religioso, Dio si manifesta nella Storia e la guida, e pur potendo concedere un qualche tipo di libero arbitrio, comunque alla fine deve ottenere la realizzazione della sua volontà; quindi la Storia ha la sua direzione, il suo fine, e l'uomo ha solo una limitata capacità d'influenza. Per questo, contrariamente allo stereotipo, molti dei maggiori teologi tendono ad essere deterministi (come Calvino) o compatibilisti (come Tommaso d'Aquino), più che libertaristi. Ma anche un filosofo ateo ha ragioni per sostenere il determinismo storico: questo Paese, la sua legge, la sua cultura, tutto esiste in virtù della sua Storia, in virtù di guerre, rivoluzioni, cambi d'ideologia e quant'altro sia stato compiuto dagli uomini del passato. Anche ammettendo che l'uomo sia stato del tutto libero ad un certo punto del tempo, comunque gli eventi di oggi sono conseguenza degli eventi passati, e dunque ci sono pesanti limiti a quali leggi si possono approvare, quali guerre si possono vincere, quali proposizioni si possono pensare.

Infine, c'è il *determinismo genetico*. La psicologia offre un ottimo metodo per stimare l'influenza della genetica sul comportamento: l'analisi statistica su gemelli separati alla nascita. Il comportamento di una persona dipende idealmente da due insiemi di variabili: ambiente (educazione, conoscenze, influenze degli amici...) e genetica. Per poter verificare cosa causa un certo comportamento devo avere controllo su queste variabili. Eliminare le differenze ambientali è impossibile: per farlo, dovrei trovare due persone che hanno condotto una vita assolutamente identica, ma casi così non esistono, neanche nei gemelli monozigoti che sono geneticamente identici... geneticamente identici? Ma allora le differenze genetiche sono eliminabili! In primo luogo, le loro differenze non possono essere imputate a fattori genetici, quindi devono dipendere da fattori ambientali. In secondo luogo, se i gemelli vengono separati alla nascita e cresciuti in famiglie diverse (idealmente in ambienti senza caratteristiche comuni rilevanti), ciò che è comune ai due non può essere imputato a fattori ambientali, quindi deve dipendere da fattori genetici. Ovviamente, al di fuori dell'idealità, vi saranno sempre alcuni elementi ambientali comuni, ma esistono tecniche statistiche per escluderli.

L'influenza di ambiente e genetica non è assoluta, bensì ha un certo peso. Se prendo coppie di persone a caso, quante volte pescherò individui di altezza molto simile? Dipenderà appunto dal caso: ipotizzo nel 30% delle osservazioni. Ma se prendo coppie di gemelli monozigoti separati alla nascita, quante volte li troverò di altezza molto simile? Spesso, direi nell'80% dei casi, e non importa se uno è stato cresciuto da una coppia di campioni di pallacanestro e l'altro da una coppia affetta da nanismo (può influire, però, l'alimentazione). Questo mi suggerisce che l'altezza è pesantemente influenzata dalla genetica. Misuro ora un altro tratto, la tendenza a sviluppare dipendenze. Se prendo coppie a caso, quante volte pesco due persone con dipendenze? Sarà raro, dico il 10%. E se prendo coppie di gemelli? Il 50% delle volte: è quindi un tratto influenzato dalla genetica, ma meno di quanto sia influenzata l'altezza.

Analisi statistiche di questo tipo sono state condotte per un gran numero di tratti: le preferenze politiche, la religiosità, vari tratti di personalità quali l'estroversione (o l'introversione), l'attrazione per materie intellettuali (o pratiche), per la novità (o la tradizione), per l'ambiguità, la capacità di seguire routine e piani con ordine, di resistere alle tentazioni, l'essere sottomesso (o dominante), empatico, emotivo, pessimista, ansioso, sono tutti aspetti influenzati dalla genetica. Vari tratti dell'intelligenza, quali la memoria, l'abilità verbale e numerica, l'abilità visiva/grafica, l'abilità uditiva/musicale, la velocità di pensiero, sono *pesantemente* influenzati dalla genetica, più dell'altezza: non importa se un gemello viene cresciuto da due premi Nobel e l'altro da due alcolisti senza diploma, è molto probabile manterranno abilità simili.

Ovviamente è possibile sollevare delle obiezioni. In primo luogo, l'ambiente è complesso, in continua evoluzione, difficile da controllare, nonostante gli sforzi per isolare l'influenza genetica è facile sottostimare sistematicamente quella ambientale. In secondo luogo, per misurare tratti come personalità e intelligenza, la psicologia deve anzitutto darne delle definizioni, le quali non sempre sono condivise: per esempio, la psicologia parla di *quoziente intellettivo*, che però molti credono non riuscire a misurare e rappresentare ciò che normalmente si intende con *intelligenza*. In terzo luogo, simulazioni statistiche mostrano che esperimenti strutturati in questo modo tendono a mostrare correlazioni anche tra gruppi di dati che in realtà sono irrelati. Ciò nonostante, i dati supportano in modo piuttosto convincente l'intuizione che alcune mie credenze, la mia capacità di adattarmi al cambiamento, di analizzare, e forse di conoscere la verità, dipendono in parte dalla mia genetica.

A differenza del determinismo nomologico, gli altri determinismi visti non sono una negazione della libertà, rimangono compatibili con l'esistenza di più futuri possibili, affermano solo che un certo futuro risulta più probabile di altri poiché, sotto tali influenze, diventa difficile (non impossibile) agire diversamente. Se il determinismo nomologico pare costituire un contesto scettico, cosa dire di questi altri determinismi?

Il linguaggio, le caratteristiche del territorio, la tecnologia, la Storia, la genetica influenzano le mie capacità di pensiero e le mie credenze. Ho ragione di credere che queste influenze siano del tutto indipendenti dalla verità: per esempio, non c'è alcuna relazione tra le caratteristiche di un territorio e la verità dei pensieri che suscita in chi lo esperisce. Potrei quindi ipotizzare che le credenze suscitate siano in ugual misura vere e false. E invece voglio imitare Putnam: i miei geni, la mia capacità di linguaggio, il modo in cui interpreto ciò che vedo, la tendenza a sviluppare tecnologia, e il mio comportamento in generale (che si traduce in eventi storici) hanno tutti origine nell'evoluzione, fanno tutt'uno con essa, e poiché l'evoluzione crea facoltà conoscitive affidabili, anche queste altre influenze devono essere tali da spingermi verso una maggioranza di credenze vere.

Questo può sembrare ancora paradossale, nello specifico sembra creare un *paradosso della prefazione*. Io credo a tutte le frasi che ho scritto in questo libro, non voglio certo lasciarvi degli errori; nella prefazione, però, ho scritto che mi prendo la responsabilità degli errori che emergeranno, poiché per esperienza so di essere fallibile: quindi, pur credendo ad *ogni* frase presa singolarmente, non credo siano *tutte* vere nell'insieme. Se io credo di essere influenzato da linguaggio, territorio, tecnologia, Storia, genetica in modo da avere una *maggioranza* ma non una *totalità* di credenze vere, generalizzo questo caso:

- credo vero che *a*, credo vero che *b*, credo vero che *c*... e via così per ogni mia credenza;
- se credo *a*, e credo *b*, allora credo anche vera la congiunzione «*a* e *b*» (è il *principio di agglomerazione*);
- quindi credo vere tutte le mie credenze congiunte;
- ma poiché sono influenzato e fallibile, credo di averne solo una maggioranza di credenze vere (se l'insieme delle credenze è abbastanza grande, è troppo improbabile siano tutte vere);
- quindi *non* credo vere tutte le mie credenze *congiunte*.

Pare vi sia incompatibilità tra il credere a ogni propria credenza (com'è inevitabile), e il credersi fallibili. Infatti, posso forse dire «credo ai protoni, credo agli elettroni, ma credo anche che uno dei due non esista»? Sarebbe assurdo, se credo vera ogni mia frase non posso negare di avere una totalità di credenze vere. Di contro, se credo di essere fallibile, e tale libro contiene un numero abbastanza alto di frasi, posso forse crederle vere tutte congiunte? Anche questo sarebbe assurdo; e poiché il principio di agglomerazione afferma che «se credo più frasi, allora credo anche nella loro congiunzione», se non credo alle mie frasi congiunte, non posso neanche crederle tutte separatamente, ve ne sarà almeno una alla quale in realtà non credo. Eppure, io davvero credo ad ogni frase del mio libro, e davvero credo che vi siano degli errori, e non percepisco nulla di problematico in ciò. Se voglio mantenere questa compatibilità, non mi resta che rifiutare il *principio di agglomerazione*: io credo *a*, credo *b*, ma posso non credere «*a* e *b*».

E com'è possibile? Se io credo veri *a* e *b*, e so che «*a* e *b*» è vero se *a* e *b* sono veri, dovrò pur credere vero anche «*a* e *b*». In effetti il principio deve rimanere valido in un gran numero di casi, ma è sufficiente che non sia necessario, cioè che sia invalido in certe rare circostanze: nello specifico, io credo tutte le frasi di questo libro, ma è la credenza nella mia fallibilità a *bloccare* l'implicazione alla credenza nella congiunzione; quindi credo le mie frasi separatamente ma non nell'insieme *perché* tra le frasi c'è la credenza nel fallibilismo (le credessi anche nell'insieme, creerei una contraddizione perché non mi riterrei fallibile).

Posto ciò, posso ora formulare un esperimento mentale.

Anzitutto, parto avendo un libero numero di credenze derivate o influenzate da linguaggio, territorio, tecnologia, Storia, genetica (posso essere, per esempio, tutte le credenze che ho io ora). Poiché queste influenze devono essere positive, immagino siano tali da garantire che due credenze su tre siano vere. Divido quindi tali credenze in tre insiemi ugualmente grandi, due con solo credenze vere, e uno con tutte le credenze false (è come dire che le credenze hanno in media il $\frac{2}{3}$ di probabilità di essere vere):

1. Vero
2. Vero
3. Falso

Io ora voglio acquisire nuove conoscenze, e posso farlo tramite il ragionamento: combinando le mie credenze attuali posso inferirne di nuove. Per semplicità, mi dò questi presupposti:

- il mio ragionamento è sempre corretto, privo di qualsiasi influenza da parte del linguaggio e degli altri fattori;
- ogni ragionamento ha due premesse (esistono ragionamenti con più premesse, ma mi limito per praticità);
- le premesse, pescate dagli insiemi di partenza, possono provenire dallo stesso insieme, oppure da insiemi diversi;
- l'ordine delle premesse è irrilevante (« x e y quindi z » e « y e x quindi z » contano come lo stesso ragionamento);
- la conclusione è vera solo se le premesse provengono da insiemi di credenze vere.

Quest'ultimo presupposto è, in realtà, sbagliato, visto che una conclusione vera può essere inferita da credenze false: ad esempio, se io credo che «Socrate è un coniglio» (falso) e che «tutti i conigli sono mortali» (vero) posso comunque concludere che «Socrate è mortale» (vero, è una nuova conoscenza). Anche in questo caso, però, voglio imitare ciò che Putnam dice sull'evoluzione: tra tutte le varie e numerose credenze false, solo una piccola percentuale sarà tale da portare comunque ad una conclusione vera; quindi è per me legittimo semplificare, immaginando che il terzo insieme contenga solo quelle credenze false che non portano a conclusioni vere.

Adesso raggruppo le conclusioni dei ragionamenti in base agli insiemi di provenienza delle premesse (che indico tra parentesi):

4. Vero (1 & 1)
5. Vero (1 & 2)
6. Vero (2 & 2)
7. Falso (1 & 3)
8. Falso (2 & 3)
9. Falso (3 & 3)

Sono partito con tre insiemi di credenze, il $33,3\%$ dei quali (cioè uno) contenente credenze false. Col ragionamento ho creato sei nuovi insiemi, dei quali il 50% è tutto falso; in totale, ho nove insiemi di credenze, false per il $44,4\%$. Ora continuo a ragionare combinando le credenze contenute in questi sei nuovi insiemi. Dover scrivere ogni combinazione a mano è troppo noioso, così ho composto questa formula per calcolare direttamente la percentuale di credenze false (e quindi il grado di inaffidabilità media delle credenze):

$$\frac{nf - \frac{1}{2}f(f-1)}{\frac{1}{2}n(n+1)}$$

dove n è il numero di insiemi di partenza, ed f è il numero di quelli falsi tra questi; il numeratore calcola il numero di nuovi insiemi falsi, il denominatore il numero totale di nuovi insiemi. Per esempio, usando questi sei nuovi insiemi (tre veri e tre falsi) scriverò: $\frac{(6*3) - (0,5*3)(3-1)}{0,5*6(6+1)} = \frac{15}{21} = 71,43\%$, con ventun insiemi dei quali quindici falsi. Posso riapplicare la formula a questi nuovi insiemi, e ottengo: $\frac{(21*15) - (0,5*15)(15-1)}{0,5*21(21+1)} = \frac{210}{231} = 90,90\%$.

C'è *propagazione dell'errore*: da un insieme di credenze affidabili ottengo credenze sempre più inaffidabili, credenze che hanno una sempre maggior probabilità di essere false; se due credenze

hanno un 50% di probabilità di essere false, la loro congiunzione (che formano quando fanno da premesse ad un ragionamento) ne avrà il 75%, e una congiunzione di congiunzioni (cioè un ragionamento le cui premesse sono le conclusioni di un altro) il 93,75%; per questo, come diceva il cardinale Newman, le conclusioni di un ragionamento sono sempre più dubbie delle sue stesse premesse, e non esistono quindi vere prove logiche. Epperò questa banalità potrebbe avere una conseguenza notevole: se io credo di essere influenzato dal linguaggio e dagli altri fattori detti, devo riconoscere che il ragionamento è un mezzo inaffidabile perché produce velocemente una maggioranza di credenze false, quindi non posso essere giustificato a credere alcuna conclusione dei miei ragionamenti, il che è indesiderabile.

Questa conclusione, però, è troppo pretenziosa. Il problema della propagazione dell'errore è un classico della statistica, e ci sono varie tecniche per gestirlo. Infatti quella che ho presentato è una caricatura del ragionamento: risulta inaffidabile perché, prendendo come premesse di un nuovo ragionamento le conclusioni di un altro senza che siano ulteriormente giustificate, ad un certo punto inizia a selezionare premesse inaffidabili; il ragionamento, però, è affidabile se ha premesse affidabili, e lo è perché, in primo luogo, io non combino tutte le premesse possibili ma seleziono le migliori, e, in secondo luogo, altri tipi di giustificazione (come test empirici, esperienze sensoriali e quant'altro) affiancano il ragionamento e aumentano l'affidabilità delle sue conclusioni.

Infine, devo sottolineare che i determinismi non-nomologici sono coerenti con l'esistenza di più futuri possibili tali che, posta una qualsiasi mia credenza, posso cessare di crederla; la propagazione dell'errore, quindi, è contingente: per tali determinismi il ragionamento non porta per forza a insiemi di credenze inaffidabili, in quanto non c'è alcuna falsità che io debba necessariamente credere e usare nei ragionamenti.

La tesi compatibilista

Oggi sono pochi i filosofi sostenitori di un puro e semplice determinismo. Sono però pochi anche i difensori del libertarismo: l'argomentazione per ragione, che prova ad individuare un'incompatibilità tra determinismo e pratica scientifica, è spesso liquidata come ingenua e banale. La grande maggioranza dei filosofi sostiene invece il compatibilismo, cioè la compatibilità del libero arbitrio col determinismo nomologico (mentre il libertarista può sostenere *solo* la compatibilità con determinismi non-nomologici).

Come mai i filosofi analitici sono così attratti da questa posizione? Principalmente perché Hume, focalizzandosi non sulla la libertà *tout court* ma su *quella libertà necessaria alla responsabilità morale*, mostra che tale interpretazione implica il determinismo stesso: infatti, se un comportamento è determinato, lo è *da una causa*, quindi un comportamento libero, in quanto indeterminato, è privo di causa e perciò imprevedibile, ma come posso essere responsabile di un comportamento libero, se non potevo prevedere il suo darsi, e io non sono la sua causa? Non posso, quindi la libertà necessita il determinismo, quindi devono essere compatibili.

I filosofi però favoriscono il compatibilismo pure per un altro motivo: collegando metafisica ed etica, questa interpretazione humeana del libero arbitrio risulta un tema rilevante anche per la teoria politica. Come scrive il filosofo Stephen Cave sull'*Atlantic*, «oggi, il presupposto del libero arbitrio attraversa ogni aspetto della politica americana», è un tema così pervasivo che perfino i giudici parlano di libero arbitrio nelle loro sentenze. Ma quindi anche un filosofo può voler difendere il libero arbitrio solo per giustificare un suo scarso interesse nel *welfare* (se ognuno è

libero chi è in difficoltà lo è per sua colpa), o al contrario può sostenere non vi sia libertà solo perché porta avanti campagne di depenalizzazione (se i criminali non sono liberi non sono neanche responsabili). In Europa, l'idea che la filosofia del libero arbitrio abbia rilevanza politica suona come una sciocchezza, qualsiasi correlazione possa esserci tra credenza nel libero arbitrio e posizione politica è sicuramente contingente e variabile nel tempo; ma tant'è.

Personalmente, non mi interessa questo compatibilismo etico, non mi interessa il legame tra libertà e responsabilità morale: i principi esposti nella sezione di etica rendono già evidenti le mie idee in merito. A me interessa solo il legame tra libertà e conoscenza: cioè, in ciò che segue, devo capire se le definizioni di *libertà* offerte dalle varie teorie sono anche definizioni di *quella libertà necessaria alla conoscenza*.

Oggi il compatibilismo richiama subito alla mente la famosissima citazione dell'anti-kantiano Schopenhauer, che affermava: «è certo che un uomo può fare ciò che vuole, ma non può volere che ciò che vuole». C'è, implicita in questa frase, una definizione di libertà: la libertà è il poter fare ciò che voglio. Avrebbe infatti senso volere la libertà di fare ciò che *non* voglio fare? Certo che no. Tale definizione è coerente col determinismo: sono determinato ad avere una certa volontà e ad agire in conformità con essa, ma sono anche libero proprio in virtù del mio *sentirmi* libero, in virtù dell'esperienza della volontarietà del mio agire. In altre parole: sono libero perché, *avessi voluto agire diversamente, avrei agito diversamente*, ma sono anche determinato perché *non posso volere diversamente*.

Questa definizione di libertà, però, non è estendibile alla libertà necessaria alla conoscenza. Ecco una versione epistemologica della citazione di Schopenhauer: «l'uomo può *credere* ciò che vuole, ma non può volere che ciò che vuole». L'unica libertà che mi è garantita è quella di credere ciò che voglio; è chiaro però che tale libertà, pur necessaria, è insufficiente alla conoscenza, come minimo devo avere anche libertà di scelta (nel senso detto discutendo il teorema di Bell). Schopenhauer, in effetti, è disposto ad offrire una definizione più larga: per lui la volontà è la *cosa in sé*, cioè è indipendente dallo spaziotempo e dalla causalità (che secondo lui si danno solo nell'esperienza), di modo che la volontà risulta libera, non determinata; ma allora questo non è più compatibilismo.

Ora prendo ispirazione dalla psicologia e modifico il compatibilismo di Schopenhauer sostituendo la volontà con un desiderio *a più livelli*. Espongo la teoria con un esempio: un tossicodipendente ha il desiderio di drogarsi, ed è determinato dalla chimica del suo cervello a desiderare ciò; ha però anche un *metadesiderio*, cioè *desidera non desiderare* di drogarsi; quando, disintossicatosi, avrà desideri in linea coi suoi metadesideri, sarà libero perché farà ciò che desidera di desiderare fare e, anche se i suoi desideri rimangono determinati, non ha senso per lui desiderare altro tipo di libertà più ampio di questo. Cioè: sono determinato se *desidero non desiderare* ciò che faccio, e sono libero se *desidero desiderare* ciò che faccio (tale è la definizione di libertà implicita nell'esempio).

Questo però non toglie validità all'obiezione già esposta: la libertà di fare ciò che desidero desiderare non è quella libertà necessaria alla conoscenza. Potrei allora aggiungere che, anche se i miei desideri (che portano all'azione) sono determinati, i miei *metadesideri* sono liberi: sono libero di *metadesiderare* ma non di agire, libero nei *metadesideri* e determinato nell'azione. Ma anche questo non è più compatibilismo: infatti mi sto limitando a dire che in me convivono sia elementi liberi che determinati, quando invece dovrei mostrare che *uno stesso elemento* può essere *contemporaneamente* libero e determinato.

Come mai questi compatibilismi non funzionano? La tesi contraria al compatibilismo, cioè l'*incompatibilismo* (che è posizione comune a libertaristi e deterministi), prevede la validità di due principi:

1. *principio delle possibilità alternative*: sono libero solo se esistono più futuri possibili;
2. *principio di originazione*: sono libero solo se sono la causa delle mie azioni.

Il compatibilismo volontarista di Schopenhauer, riducendo la libertà alla volontà (o, nella seconda versione, al desiderio), finisce per affermare che sono libero *se mi sembra di essere tale*; in questo modo, può negare entrambi i principi: posso essere libero anche se non esistono più futuri né sono causa delle mie azioni, è sufficiente che io *abbia l'esperienza* (pur illusoria) di compiere scelte. Così però ne risulta una definizione di libertà troppo controintuitiva: questo compatibilismo è, come diceva Kant, solo un gioco di parole, chiama «libertà» qualcosa che non lo è perché troppo diverso da ciò che normalmente si indica con tale nome.

Il compatibilista, quindi, non può negare entrambi i principi, ma non può neanche accettarli entrambi: infatti la loro combinazione crea una definizione di libertà incompatibile col determinismo. Non resta che negarne uno, e così si formano le due principali teorie compatibiliste:

- se nego le *possibilità alternative* ma mantengo l'*originazione*, dirò che sono libero *perché* sono causa delle mie azioni, *anche se* sono determinato perché esiste un unico futuro possibile;
- se nego l'*originazione* ma mantengo le *possibilità alternative*, dirò che sono libero *perché* esistono più futuri possibili, *anche se* sono determinato perché non sono io la causa delle mie azioni.

L'esistenza di più futuri possibili è necessaria alla libertà 1: i casi di Frankfurt e altri esperimenti mentali

Parto dalla prima: originazione senza possibilità alternative. Frankfurt è il più grande sostenitore di tale teoria: anche se è più noto per la sua raffinata analisi del concetto di *stronzata*, la sua influenza nel campo del libero arbitrio è tale che le argomentazioni in difesa di questo tipo di compatibilismo sono chiamate direttamente *casi di Frankfurt*. Eccone uno.

Si tengono le elezioni, e sono indeciso tra il candidato di destra e quello di sinistra. A mia insaputa, quello di destra (che chiamerò con un nome di fantasia, Matteo) mi ha fatto impiantare un chip nel cervello: se io scelgo di votare a destra, Matteo non fa niente; se però scelgo di votare a sinistra, Matteo attiva questo chip, il quale mi fa subito cambiare idea e mi spinge a votare a destra. In questo contesto, sono libero di scegliere chi votare?

Rispongo l'esempio in modo meno fantascientifico: Matteo è molto bravo nel valutare le intenzioni altrui, ed è anche molto bravo nell'oratoria, tanto che riesce sempre a persuadere coloro a cui parla; al che, vedendomi, se Matteo capisce che voglio votare a destra, non farà alcunché; se però vede che voglio votare a sinistra, mi persuaderà con la sua arte in modo tale da farmi scegliere infallibilmente di votare a destra. Matteo agisce in modo meccanico: ad una mia scelta segue una sua reazione che è l'unica possibile in quel caso. Ma io, sono libero nel mio votare?

Io ho due volontà possibili: sono libero di scegliere se votare a destra, o a sinistra. Se scelgo di votare a destra, non essendo soggetto ad influenze esterne, sono ben disposto a dire di essere libero: lo sono perché sono io la causa della mia scelta, e tale scelta discende da una volontà che è libera. Epperò, in quanto al voto, non esistono più futuri possibili: non posso agire diversamente da come agisco, è per me impossibile votare a sinistra in quanto entrambe le mie volontà possibili si risolvono nell'unico futuro possibile del voto a destra. Quindi devo ammettere che il principio delle possibilità alternative è inessenziale alla libertà, è sufficiente il principio di originazione: se io

agisco in un certo modo *perché* quella è l'unica scelta, non mi dico libero, ma se agisco in quel modo perché ho mie motivazioni, allora, *anche se* è l'unica scelta, sono comunque libero.

In sostanza: l'argomentazione, del tutto intuitiva, descrive un caso tale che, pur essendovi un unico futuro possibile, sono disposto a definire «libero», implicando che il senso che assegno a questa parola coincide col solo principio di originazione. Sarà ciò vero anche per quella libertà necessaria alla conoscenza? Devo rivisitare l'argomentazione in senso epistemologico.

Al che: con ancora quel chip nel cervello, voglio capire se la Terra sia o meno piatta come un vinile; se io trovo un motivo (foss'anche solo pragmatico) per giustificare l'idea che la Terra sia tonda, il chip non fa niente; se però trovo un motivo per credere che la Terra sia piatta, il chip modifica l'attività del mio cervello in modo tale che io creda lo stesso nella sua rotondità. In tale contesto, posso acquisire conoscenza sulla forma della Terra?

Seguendo le conclusioni del precedente esempio, dovrei dire: sono libero di credere che la Terra sia tonda (perché l'ho scelto), ma non di non crederci (perché non potevo scegliere altro). Nell'argomentazione per ragione, però, ho detto che, se credo ciò che credo a prescindere dal suo valore di verità, allora la mia credenza non può essere indefettata. In questo caso, anche se io ho ragioni per credere che la Terra sia tonda, il contesto è tale che, fosse piatta, lo crederei comunque: perciò, sapessi di avere un chip nel cervello che mi porterebbe sempre e comunque a credere la stessa cosa, non mi riconoscerei giustificato (lo sarei, invece, se fosse rotto o mai attivato). Quindi il tipo di libertà descritto è insufficiente alla conoscenza.

Voglio continuare ad esplorare le mie intuizioni tramite un esempio personale.

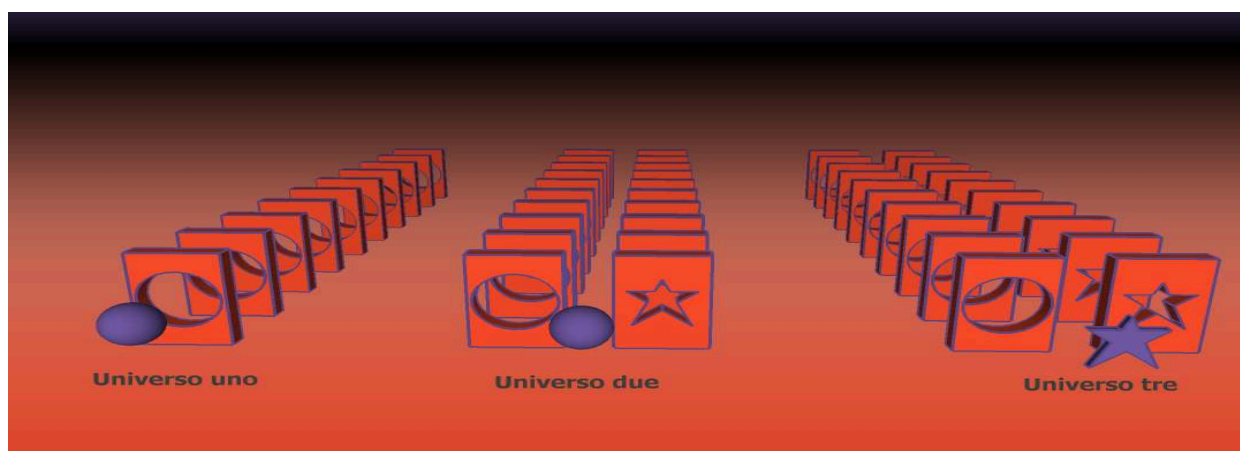


Figura 14

Ci sono tre universi-modello semplicissimi, tutti composti da due soli elementi: un solido e delle porte. In tutti e tre vigono solo queste leggi fisiche:

- ad ogni istante, il solido deve attraversare la porta a sé più vicina;
- il solido non può attraversare più volte la stessa porta;
- se due porte sono equidistanti, il solido sceglie quale attraversare.

Ovviamente, il solido rappresenta un qualsiasi soggetto capace di compiere scelte, e le porte rappresentano gli stati futuri dell'universo; le tre leggi quindi rappresentano, in ordine,

l'impossibilità di fermare il tempo, l'impossibilità di viaggiare indietro nel tempo, e la possibilità di scegliere tra più futuri possibili.

Nel primo universo-modello il solido è di forma sferica, e ha davanti un'unica fila di porte, tutte con un'apertura circolare abbastanza grande da lasciarlo passare. In tale universo, quindi, esiste un unico percorso possibile: ad ogni istante il solido deve attraversare la porta che ha di fronte, e non può fare altro. Questo modello rappresenta gli universi possibili nomologici, cioè con un unico futuro possibile: poiché ad ogni stato ne può seguire uno solo, è impossibile compiere scelte. Di conseguenza, in un universo di questo tipo non c'è libertà né conoscenza per i motivi che ho discusso nell'argomentazione per ragione.

Nel secondo universo-modello il solido è di nuovo una sfera, ma ha davanti due file di porte: la prima fila è composta tutta da porte con un'apertura circolare abbastanza grande da lasciar passare la sfera; la seconda, invece, è composta tutta da porte con un'apertura a forma di stella attraverso il quale non può passare. Poiché le prime porte di ogni fila sono equidistanti dalla sfera, questa può scegliere quale attraversare; è però chiaro che la porta con l'apertura circolare costituisce una *scelta obbligata*, essendo l'unica idonea alla forma del solido.

Questo modello rappresenta gli universi possibili immaginati dai *casi di Frankfurt*. Dico che un futuro mi è *accessibile* se posso agire in modo tale da finire per viverlo, cioè se le mie scelte sono una condizione del darsi di quel futuro. Orbene, questi casi sembrano offrire più futuri possibili (mi sembra di poter votare a destra come a sinistra), ma solo uno di questi futuri è accessibile (perché in nessun caso posso finire per votare a sinistra). Dirò così che la prima fila rappresenta un futuro *reale*, cioè che può realizzarsi e nel quale posso vivere; la seconda rappresenta invece un futuro *virtuale*, cioè che esiste come possibilità ma che è *come non esistesse* in quanto in nessun caso posso realizzare le condizioni affinché si realizzi, è una possibilità logica ma un'impossibilità metafisica.

Mi chiedo: in questo caso, sono disposto ad assegnare un qualche tipo di libertà di scelta al solido? L'intuito mi dice di no, perché non ci sono più futuri che possono *davvero* darsi, non c'è una concreta possibilità di scelta in quanto si darà un certo futuro qualsiasi cosa possa fare; quindi, per motivi già esposti, l'esistenza di futuri virtuali è insufficiente a quella libertà necessaria alla conoscenza, servono almeno due futuri reali.

Nel terzo universo-modello il solido è a forma di stella, e ha di fronte due file di porte equidistanti: la prima è composta da porte con un'apertura circolare, la seconda da porte con un'apertura a forma di stella, ed in entrambi i casi l'apertura è abbastanza grande da lasciarlo passare. Questo modello rappresenta quindi gli universi possibili con più futuri reali, e non ho motivo di negare la libertà del solido.

Ora però voglio dare a questo solido una nuova caratteristica, che chiamo *razionalità*: cioè, quando deve scegliere quale porta attraversare, poiché vuole evitare di colpire gli spigoli per errore, sceglierà sempre quella più spaziosa, che in questo caso è quella circolare. Quindi, il solido *potrebbe* scegliere le porte della seconda fila, ma *di fatto* sceglierà sempre e solo quelle della prima.

Il modello rappresenta ora un universo con determinismo non-nomologico: ci sono più futuri possibili, ma poiché io sono perfettamente razionale, e solo uno di questi futuri possibili è razionalmente desiderabile, pare che io sia determinato dalla mia stessa razionalità. Mi chiedo: questo rende uno dei due futuri *virtuale*, e distrugge quindi la libertà di conoscere?

Credo non sia così. Nel secondo universo, l'inaccessibilità del futuro virtuale è *indipendente* dalla scelta, è un futuro che non può realizzarsi neanche a volerlo; nel terzo universo, invece, l'inaccessibilità del futuro reale è *dipendente* dalla scelta: è vero che *manifestando razionalità* posso accedere a solo uno dei due futuri possibili, ma poiché la razionalità è una mia proprietà contingente, posso anche decidere di non manifestarla, posso scegliere di agire irrazionalmente, e così accedere all'altro futuro secondo la mia volontà.

Emerge così un fatto interessante: nei tre universi, il solido si comporta sempre nello stesso modo, cioè attraversa solo le porte con apertura circolare; ciò nonostante, in uno di questi universi il solido è libero, e negli altri no. Addirittura, posso immaginare esista una mia controparte identica a me in tutto escluso il fatto che il suo comportamento è determinato, e io, proprio perché sono libero, posso scegliere di agire sempre in modo identico a questa mia controparte (pur avendo la possibilità di distinguermi, di agire anche in altri modi). La libertà, quindi, non è individuabile nel comportamento: io potrei essere libero ma agire *come fossi* del tutto determinato; quindi il fatto che io agisca in tutto e per tutto in linea con le leggi fisiche, cioè che tutto il mio comportamento sia potenzialmente descrivibile da una formula matematica, non implica che io non sia libero.

Per questo motivo, i *casì di Frankfurt* sono da rifiutare. Infatti, se la libertà è riducibile al principio di originazione (senza possibilità alternative), allora affinché io sia libero è sufficiente che agisca *come agirei se vi fossero più futuri reali*. Epperò, se io sono libero e ho compiuto una serie di scelte, ed esiste un universo possibile con un unico futuro composto da tutte e sole le scelte che ho compiuto, questo rimarrebbe un universo del primo tipo dove nessun ente è libero; quindi, se un ente è determinato, il fatto che sia determinato a fare proprio ciò che fa un ente libero è insufficiente a renderlo libero a sua volta. Similmente, il fatto che io sia determinato a credere proprio ciò che crederei se fossi un genio dotato di libero arbitrio è di per sé insufficiente alla giustificazione (infatti non posso essere giustificato a credere che crederei ciò che credo anche se fossi libero). Perciò libertà e conoscenza richiedono il principio delle possibilità alternative.

L'esistenza di più futuri possibili è necessaria alla libertà 2: l'argomentazione di Van Inwagen

Come C. S. Lewis, van Inwagen è un importante filosofo della religione convertitosi al cristianesimo in età adulta; e come lui, è autore di quella che è considerata la più potente argomentazione incompatibilista. Invero, il merito di van Inwagen sta nell'aver espresso in modo rigoroso, tramite logica modale, un'intuizione ovvia, la prima banale obiezione che chiunque solleva quando inizia a riflettere sulla possibilità del compatibilismo. Rendo ora tale argomentazione in lingua naturale:

1. se un fatto è necessario, allora nessuno poteva fare alcunché per impedire il suo darsi, cioè è *inevitabile*;
2. se un fatto A è inevitabile, ed è inevitabile che se si dà il fatto A allora si darà il fatto B , anche il fatto B è inevitabile (per esempio, se il fulmine è inevitabile, ed è inevitabile che al fulmine consegua il tuono, anche il tuono è inevitabile);
3. per il determinismo nomologico, è necessario che, se si danno certi eventi passati e certe leggi fisiche, allora si darà un certo evento futuro (per esempio, il fatto che «se ha piovuto e valgono le leggi fisiche di questo universo, allora la strada sarà bagnata» è necessario perché, essendovi un unico futuro possibile, è impossibile che alla pioggia segua un futuro diverso da quello della strada bagnata);

4. quindi è necessario che, se si dànno certi eventi passati, allora si dia che, se si dànno certe leggi fisiche, si darà un certo evento futuro (dal punto di vista logico, questa frase è equivalente a quella del punto tre: se ha piovuto allora si dà che, se valgono le leggi fisiche di questo universo, allora la strada sarà bagnata);
5. quindi, per il punto uno, è inevitabile che, se si dànno certi eventi passati, allora si dia che, se si dànno certe leggi fisiche, si darà un certo evento futuro;
6. gli eventi passati sono imm modificabili, quindi necessari, e quindi inevitabili;
7. quindi, per il punto due, è inevitabile che, se si dànno certe leggi fisiche, allora si darà un certo evento futuro;
8. ma anche le leggi fisiche sono imm modificabili, quindi necessarie, quindi inevitabili;
9. quindi, per il punto due, l'evento futuro è inevitabile;
10. il libertarismo nega l'inevitabilità del futuro, il determinismo la implica, quindi le due posizioni sono incompatibili.

Come l'argomentazione per ragione, anche questa è stata criticata in tutti i modi, e come C. S. Lewis, anche van Inwagen ha riconosciuto la necessità di modificare questo ragionamento troppo tortuoso. La sua versione più semplice e diretta si deve a Finch e Warfield:

1. se un fatto A è inevitabile, ed è necessario che se si dà il fatto A allora si darà il fatto B , anche il fatto B è inevitabile;
2. è necessario che, se si dànno certi eventi passati e certe leggi fisiche, allora si darà un certo evento futuro (questa è la tesi determinista);
3. ma gli eventi passati e le leggi fisiche sono inevitabili;
4. quindi l'evento futuro è inevitabile.
5. il libertarismo nega l'inevitabilità del futuro, il determinismo la implica, quindi le due posizioni sono incompatibili.

Per van Inwagen il compatibilista deve sostenere contemporaneamente la tesi libertarista (cioè l'evitabilità del futuro) e la tesi determinista (cioè che, dato il passato e le leggi fisiche, il futuro segue necessariamente). L'argomentazione, però, mostra che le due tesi sono incompatibili: il libertarista deve credere che esistano più futuri possibili reali, il determinismo implica invece un unico futuro reale. Quale punto del ragionamento può rifiutare il compatibilista per rendere le due tesi coerenti? Poiché il primo punto pare logico, e senza il secondo non c'è compatibilismo, non resta che rifiutare il terzo: il compatibilista deve sostenere che il passato e le leggi fisiche siano evitabili, cioè che l'azione umana possa modificare il passato e la fisica: così resta vero che un certo passato implica un unico futuro possibile, ma ci sono più futuri possibili reali perché il passato può cambiare. Questa idea, però, è assurda: nessuno può cambiare il passato.

David Kellogg Lewis, però, obietta che questo ritratto delle credenze compatibiliste pecca di carità. Secondo lui, il compatibilista, pur credendo in un unico futuro reale, non crede affatto che *avessi scelto di agire diversamente* allora *avrei causato un cambiamento* nel passato e nelle leggi fisiche; crede invece, ribaltando le componenti della frase, che *se il passato e le leggi fisiche fossero stati diversi* allora *avrei agito diversamente*. David Lewis, cioè, riconosce che la libertà è *la possibilità di fare altrimenti*, è l'esistenza di mondi possibili nei quali agisco diversamente rispetto al mondo attuale, ma secondo lui ciò non richiede l'esistenza di più futuri possibili reali dato uno stesso passato: io sono libero perché avrei davvero potuto fare altrimenti, *se passato e leggi fisiche fossero stati diversi*; esistono più mondi possibili e quindi futuri diversi virtuali, ma non esistono mondi possibili che, avendo un identico passato, hanno un diverso futuro.

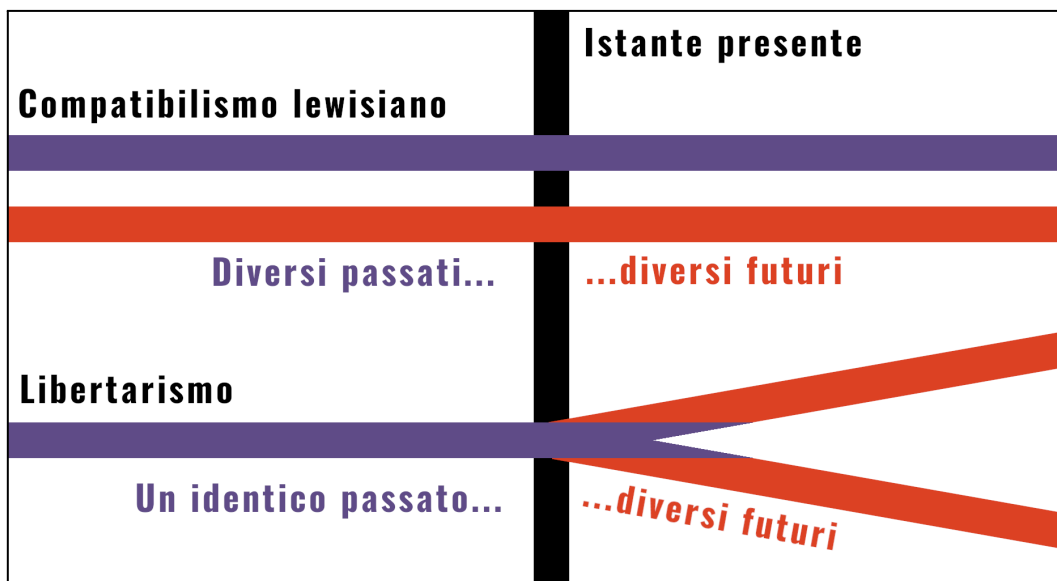


Figura 15

David Lewis ha ragione: il compatibilista non crede all'assurdità, non crede che l'uomo abbia il potere di modificare il passato; non lo crede, però, perché adotta una definizione di libertà che non richiede l'evitabilità del futuro: io sono libero perché esistono diversi futuri possibili (come chiede il libertarista) ma ad ogni passato corrisponde un unico futuro, quindi c'è un unico futuro reale e molti virtuali (come vuole il determinista). Epperò, per il libertarista questa nozione di libertà è insufficiente, il libertarista vuole più futuri reali *anche con un identico passato*, e quindi pure il compatibilista, per essere tale, *dovrebbe* credere a delle assurdità perché *dovrebbe* riconoscere, per i motivi esposti finora, che la libertà è tale se il futuro è evitabile.

Posso immaginare anche un compatibilismo *anti-lewisiano*: come vuole il libertarista, esistono più futuri possibili reali *anche dato un unico passato* e, come vuole il determinista, dato un certo passato segue necessariamente *un certo futuro*; questo *certo futuro*, però, è descritto non da un unico stato dell'universo, ma da una congiunzione di stati, cioè dato il passato *A* vi seguirà necessariamente il futuro «*B o C*», se c'è un fulmine allora necessariamente «o vi sarà il tuono o non vi sarà». Ma questa è una errata caratterizzazione del determinismo: il determinista crede che il passato implichi necessariamente un *unico* stato futuro, e non una disgiunzione di stati, quindi neanche questo tipo di compatibilismo è accettabile.

L'esistenza di più futuri possibili è necessaria alla libertà 3: ulteriori conseguenze

Oltre a rifiutare i casi di Frankfurt e a supportare l'argomentazione di van Inwagen, la necessità di avere più futuri reali rifiuta anche l'idea che la casualità sia sufficiente alla libertà. Infatti un futuro è *reale* solo se ci sono condizioni, da me realizzabili, alle quali si sarebbe dato un futuro diverso; ma se, posti più futuri possibili, questi si danno totalmente a caso, non esistono particolari condizioni, nessun evento è sufficiente al loro darsi (qualsiasi cosa io faccia, potrebbe darsi un certo futuro come un qualsiasi altro) e quindi non ci sono neanche più futuri reali: mi ritroverei sempre e comunque nel futuro *di cui si dà il caso fortuito*, perciò, già nel momento in cui agisco, *dato un certo caso fortuito*, solo un futuro possibile mi risulta accessibile (quello che *verrà* scelto dal caso).

Un'obiezione intuitiva all'esistenza di più futuri reali sta nel fatto che, fosse necessaria alla libertà, porterebbe ad una contraddizione. Sono dal gelataio, indeciso tra il gelato alla graviola e quello al mangostano; se io avessi un motivo per scegliere quello alla graviola, allora questa motivazione mi determinerebbe, esisterebbe solo il futuro nel quale scelgo la graviola (scegliessi quello al mangostano, sarei in realtà motivato a scegliere quello); quindi, se ci sono più futuri reali, non posso avere motivazioni, ma se non ho motivazioni allora agisco a caso, e quindi non sono libero: dunque, poiché io o ho motivazioni (e sono determinato) o non le ho (e agisco a caso), la libertà è impossibile.

Questo mi ricorda di un'antica diatriba teologica. Dio è l'essere perfettissimo, quindi è totalmente libero; eppure, proprio perché perfettissimo, deve creare un mondo perfetto: infatti, scegliesse di non creare alcunché, sarebbe imperfetto. Ma, se non può non creare, allora non è libero, quindi è imperfetto... paradossale, un ente perfetto deve essere libero, ma se è perfetto non può essere libero. Come risolvere?

Sia l'esempio del gelato, sia questo paradosso teologico, si basano sullo stesso equivoco già discusso in merito alla razionalità: come la razionalità non è una causa sufficiente all'azione (in quanto posso anche agire irrazionalmente), così la motivazione a scegliere un certo gusto di gelato è una causa necessaria ma non sufficiente alla scelta, in quanto mantengo la libertà di scegliere contro le mie stesse motivazioni (anche se magari, di fatto, non lo farò mai); e così Dio, come dice Leibniz, essendo perfettissimo, ha scelto di creare, ma poteva pure astenersi dal farlo (anche se, per sua natura, non avrebbe mai scelto così).

Ora generalizzo al massimo l'esempio del gelato. Quando io osservo dei fenomeni, o scopro in essi una legge, un *pattern*, una funzione che collega lo stato presente dell'universo al successivo, oppure non vi trovo alcuna ricorrenza del genere: nel primo caso, il fenomeno è determinato da una legge, mentre nel secondo, non seguendone alcuna, è casuale. Ovviamente o posso individuare una legge o non posso, quindi un evento o è determinato o è casuale, e la libertà, essendo incompatibile con entrambe le opzioni, risulta impossibile. Dov'è l'errore questa volta?

Sta nell'essermi dimenticato di distinguere l'*apparire* dall'*essere*. Un ente libero, in quanto tale, è appunto libero di agire seguendo una legge, ma anche libero di agire a caso, a seconda delle sue scelte. Quindi il comportamento di un ente può essere *descritto* solo come determinato o casuale, ma questo non vuol dire che possa *essere* solo determinato o casuale, potrebbe anche *essere libero e apparire determinato, apparire casuale*. Perciò un ente può *essere* libero, ma non può anche *apparire* libero: il fatto che un atomo risponda a certe leggi matematiche, o che decada a caso, nulla mi dice sulla sua libertà.

L'originazione è necessaria alla libertà

Il principio delle possibilità alternative pare essenziale alla libertà, e se l'universo fisico risulta essere indeterminato, cioè se davvero esistono più futuri possibili, è anche soddisfabile. Il compatibilista, quindi, può solo rifiutare il principio di originazione: così, definisce la libertà come l'esistenza di più futuri (indipendentemente da quale sia la causa del comportamento), e la determinatezza come lo stato nel quale il proprio comportamento ha una causa a sé esterna (indipendentemente da quanti futuri possibili esistano). È questa una tesi difendibile?

Sulla base di quanto detto sopra, il principio delle possibilità alternative con futuri reali (non virtuali) implica che:

- se la mia azione è condizione sufficiente al darsi di un certo futuro, e nessun altro evento antecedente o contemporaneo alla mia azione lo è, sono libero, in quanto al momento dell'azione esistono più futuri possibili reali che posso selezionare;
- se gli eventi antecedenti o contemporanei alla mia azione sono sufficienti al darsi di un certo futuro, allora, sia che la mia azione sia sufficiente sia che non lo sia, io non sono libero: infatti, quando agisco, non esistono già più futuri alternativi a quello selezionato dagli eventi passati;
- se ogni evento è insufficiente al darsi di un certo futuro, e lo sviluppo del mondo è quindi del tutto casuale, io comunque non sono libero, per i motivi detti.

In altre parole: è necessario che il mio agire sia determinante, che sia rilevante nel determinare il darsi o non darsi dei futuri possibili. Devo quindi affermare un *principio di controllo* metafisico: io sono libero, relativamente ad un certo futuro, se ho controllo su di esso, cioè se il suo darsi è in funzione delle mie azioni e nessun altro evento antecedente o contemporaneo gli è sufficiente.

Faccio un esempio. Ho una pistola: premo il grilletto, che causa lo sparo dei proiettili, che causa la morte del Presidente. Anzitutto, chi è che l'ha ucciso? Sarebbe assurdo dire, come Clint Eastwood, «non l'ho ucciso io, l'hanno ucciso i miei proiettili»: sono proprio io l'assassino perché, se *A* causa *B* che causa *C*, e *B* causa *C* perché senza il primo non si dà il secondo, anche *A* causa *C*, poiché senza *A* non si dà né *B* né *C*. *C*'è una differenza, però, tra me e i proiettili. I proiettili causano la morte del Presidente *senza avere controllo* in quanto il futuro possibile nel quale il Presidente non muore è, dalla loro prospettiva, *virtuale*: posta la mia scelta di premere il grilletto, i proiettili possono solo colpire e uccidere, cioè la mia scelta è sufficiente a determinare la morte futura, e un futuro diverso potrebbe darsi solo con un diverso passato. Io, invece, causo la morte del Presidente *avendo controllo*: indipendentemente da tutti gli altri eventi, potevo decidere di non premere il grilletto, quindi di non sparare, quindi di non uccidere, perciò il futuro nel quale il Presidente non muore era per me *reale*.

Io ho controllo sulla morte del Presidente perché ho controllo sullo sparo che la causa; di contro, non avessi controllo sulla causa, non avrei neanche controllo sulla morte. Per esempio, se io decido di uccidere e punto la pistola, ma la pressione del grilletto è causata da un tic sul quale non ho controllo, non posso neanche dire di avere controllo sulla morte del Presidente in quanto, una volta puntata la pistola, il darsi del tic è sufficiente all'uccisione, ma questo suo darsi o non darsi non è in funzione delle mie azioni.

Lo stesso vale per gli eventi che causano la mia decisione. Immagino che un mago mi abbia colpito con un incantesimo tale da permettergli di controllare tutte le mie decisioni. Il principio delle possibilità alternative rimane valido perché, essendo il mago libero, può decidere sia di farmi sparare, sia di non farmelo fare; ma poiché io devo fare ciò che lui deciderà, queste possibilità esistono *per lui*, non *per me*, è lui ad avere controllo e non io in quanto, data la sua scelta, non posso più agire in altro modo. Io potrei avere controllo sullo sparo, ma solo se avessi controllo sulle decisioni del mago; se non lo ho, è come se io stesso fossi un proiettile.

L'essere causa di un evento è diverso dall'averne controllo, ma è chiaro che io posso avere controllo su un evento solo se ho un qualche rapporto causale con esso: se le mie azioni non figurano nella catena causale che porta a quell'evento, non posso neppure influire sul suo darsi o non darsi; quindi, se io controllo un evento, allora o sono la sua causa, o causo ciò che lo causa.

Questo vuol dire che il principio di controllo implica il principio di originazione: poiché io sono libero solo se ho controllo, sono libero solo se sono causa delle mie azioni. Di conseguenza, anche questo secondo compatibilismo è da rifiutare in quanto la libertà, se richiede possibilità alternative reali, allora richiede anche originazione.

Il compatibilista ora, per difendere la sua tesi, deve rifiutare il principio di controllo: un'impresa ardua, visto che molti filosofi lo difendono come intuitivo. Epperò, è possibile mostrare come il suo rifiuto sia desiderabile; infatti, tale principio pare addirittura rendere la libertà impossibile: per esempio, il mio comportamento è causato dagli eventi elettrochimici del mio cervello, sui quali non ho controllo; oppure, di certo non ho controllo sull'esistenza dell'ambiente, la quale però è causa delle mie esperienze e dei miei desideri (per esempio, il mio voler bagnare una pianta vedendola secca); di conseguenza, pare che io non possa avere controllo su alcuna azione.

L'obiezione è corretta: se io sono determinato dagli eventi elettrochimici del mio cervello o dall'ambiente al quale rispondo causalmente, cioè se dati questi fatti esiste per me un unico futuro possibile, io davvero non ho controllo su alcuna azione. Ma se io sono libero, allora non sono totalmente determinato da queste cose: il vedere la pianta secca fa sorgere in me il desiderio di bagnarla, e così mi influenza, ma non mi determina a farlo, posso anche lasciarla morire di sete in quanto la pianta è una causa necessaria ma non sufficiente al mio darle da bere; quindi, è vero che non ho controllo sui desideri che l'ambiente fa nascere in me (il loro sorgere può essere determinato, non tutto in me deve essere libero), ma comunque io continuo ad avere controllo sulle mie azioni.

Il filosofo Alfred Mele, discutendo il rapporto tra fortuna e libertà, propone un esperimento mentale più raffinato. Lo scienziato Pazzo e il suo collega Genio sono entrambi innamorati della scienziata Calcestruzia. Pazzo cede alla gelosia e decide di fare fuori Genio, ma con calma: invece di accoltellarlo o avvelenarlo, Pazzo modifica il genoma di uno zigote in modo tale che, una volta nato, sarà destinato a sviluppare la volontà di uccidere Genio e a fare tutto il necessario per raggiungere tale scopo. Intanto altrove, per pura coincidenza, si sviluppa in modo del tutto naturale uno zigote identico a quello creato da Pazzo, e anche questo svilupperà la volontà di uccidere Genio. Nel caso dello zigote geneticamente modificato, la causa della sua volontà è il lavoro di Pazzo, ma mancando di controllo su tale evento lo zigote non controlla neanche le sue azioni; nel caso dell'altro zigote, invece, la sua volontà non ha alcuna causa esterna, quindi nulla impedisce il controllo sulle azioni. Ma perché dovrei dire che uno zigote è libero, e l'altro no? Poiché i due zigoti sono identici, e la loro origine (se naturale o *in vitro*) pare irrilevante, o sono entrambi liberi, o nessuno dei due lo è; di conseguenza, l'aver controllo sulle cause del proprio agire è inessenziale alla libertà.

Le obiezioni a tale esperimento mentale sono però quelle già viste: se la volontà di uccidere è *irresistibile*, allora nessuno dei due zigoti sarà libero, poiché nessuno dei due ha controllo sui processi (naturali o artificiali che siano) che hanno formato il loro genoma; di contro, se la volontà è *resistibile*, allora entrambi gli zigoti sono liberi perché, anche se le modifiche genetiche apportate influenzano le volontà di uno, comunque entrambi possono agire anche contro la propria volontà.

Ma c'è ancora un'obiezione al principio di controllo. Se un certo fatto causa le mie azioni, le quali causano a loro volta un certo evento, io posso controllare questo evento solo se controllo anche la causa delle mie azioni. Non posso però risalire infinitamente: è ovvio che non ho controllo sulla mia stessa nascita, e men che meno sulla comparsa dell'universo. Quindi, affinché io controlli un evento, le mie azioni devono causarlo (o causare la sua causa), ma non possono essere causate a loro volta da alcunché al di fuori del mio controllo; cioè, le mie azioni devono essere *nuove*, devono

nascere con me ed essere slegate dalle catene casuali degli eventi che mi precedono. In altre parole, io ho controllo solo se sono *causa prima*, *causa incausata*, solo se, come nota Chisholm, io ho con gli eventi lo stesso rapporto che Dio ha con l'universo. Ma, come anche van Inwagen teme, questo potrebbe rendere il concetto di libero arbitrio incoerente: infatti, se le mie scelte libere non sono causate da alcunché, non dovrebbero essere casuali?

Ciò che distingue la *causa prima* dalle *cause ordinarie* non è la sua precedenza temporale, ma la sua *indipendenza*. Se io sono *causa prima* allora possono esistere più mondi possibili identici in tutto tranne che per una mia azione, cioè le mie azioni devono essere autonome, indipendenti da ogni caratteristica dell'universo: finché non agisco, non esiste alcun fatto dal quale implicare come agirò. Ma questo, per la distinzione tra apparire ed essere, non significa che io debba agire a caso: posso comunque avere motivazioni e preferenze; anzi, poiché preferisco qualcosa *su qualcos'altro*, è il concetto stesso di preferenza a richiedere più futuri ugualmente possibili.

Il concetto di *causa prima* così esposto pare coerente, ma permane un enigma: se, qualsiasi motivazione abbia, io posso comunque agire in modi diversi, cos'è che mi porta a scegliere come agire, se non il caso? Devo rispondere «io», o meglio «la mia volontà»: io *esercito* volontà scegliendo cosa fare, e anche scegliendo se sceglierlo a caso o seguendo un criterio. È possibile, anche se non necessario, che questa volontà sia atemporale (esempio teologico: Dio può essere causa prima di un universo eterno, con catene casuali infinitamente estese nel passato, perché almeno in parte al di fuori del tempo). Non voglio però tentare di analizzare ulteriormente la natura della volontà: infatti è un tema comune nella tradizione continentale, ma l'analitica preferisce delegarlo alla psicologia.

È possibile che questa concezione di libertà sia troppo forte, che richieda logiche particolari, e che la scienza attuale non abbia i mezzi per poterne rendere merito. Ma tutto ciò è secondario: questa è libertà in senso pieno, ed è necessaria per la conoscenza, quindi se voglio rifiutare lo scetticismo devo accettarla, come sia possibile accordarla con la scienza è un problema successivo.

Detto tra parentesi: le leggi fisiche potrebbero essere apparenti

Come Spinoza prima di lui, Hume, pur compatibilista dal punto di vista morale, credeva che il libero arbitrio fosse un'illusione dovuta al mio essere conscio dell'azione ma non delle sue cause: quando un cane vede un giocattolo elettronico muoversi da solo, gli abbaia contro e così mostra di considerarlo cosciente e libero, ma conoscesse la causa del movimento non lo considererebbe tale; se io mi convinco di poter muovere le foglie degli alberi con la telecinesi, percepirò i loro movimenti come in accordo con la mia volontà e quindi liberi, ma non li percepirei tali se sapessi che in realtà sono causati dal vento.

Ciò nonostante, la filosofia di Hume non è in sé ostile al libertarismo. Parto dalla sua teoria della causalità, per la quale io posso osservare successioni di eventi, ma non rapporti causali in astratto: quando dico che un evento causa l'altro intendo solo dire che, qualsiasi siano le circostanze, ogni volta che osservo uno osserverò anche l'altro. Ma, per il *problema dell'induzione*, io non potrò mai osservare ogni caso possibile: io dunque posso individuare al massimo *leggi apparenti*, cioè posso descrivere le *ricorrenze non necessarie* che caratterizzano il mondo ma non pongono limiti agli eventi futuri, a quanto ne so domani potrei svegliarmi e scoprire che l'universo ha preso a comportarsi secondo una nuova fisica. Dall'esistenza di tali ricorrenze, finché non individuo eccezioni, posso abduire l'esistenza anche di *leggi metafisiche necessarie*, cioè posso affermare che in effetti ci sono leggi, dotate di esistenza concreta, che creano le regolarità e determinano un unico futuro

possibile; epperò posso anche sostenere quella che chiamo *teoria humeana delle leggi di natura* (anche se Hume non l'accetterebbe in questi termini), ovvero sia l'idea che non esistono leggi metafisiche, sono io che creo leggi apparenti individuando le ricorrenze nella mia esperienza. E questo è ben compatibile con il libertarismo.

Un'idea simile la trovo anche in David Kellogg Lewis che, essendo metafisico di scuola australiana, esprime il suo pensiero tramite uno stravagante recupero di categorie antiche, ma sempre difendendo il naturalismo scientifico e il senso comune. La tesi centrale e unificante di tutto il suo lavoro è la *supervenienza humeana*: tutto ciò che esiste consiste solo in una certa distribuzione di proprietà naturali nello spaziotempo, e da questa distribuzione si implica tutto il resto (quindi, per esempio, se due mondi possibili hanno un'identica distribuzione di proprietà naturali, hanno anche identiche proprietà etiche non-naturali). Questo però significa anche che le leggi causali non hanno esistenza autonoma: la causalità *superviene* sulla distribuzione delle proprietà, gli enti si comportano in un certo modo per loro natura e non perché vincolati da leggi, se voglio cambiare il comportamento di un ente non posso modificare fantomatiche leggi metafisiche senza modificare l'ente stesso. In quanto eternalista, David Lewis crede comunque in un universo determinato: gli eventi non si causano l'un l'altro, ma sono già distribuiti in ogni tempo. Nulla però vieta di reinterpretare la teoria in un'ottica presentista, e in tal caso risulta compatibile con il libertarismo.

Devo quindi chiedermi: sono le leggi fisiche a creare questo mondo, o sono invece i fatti di questo mondo, come il mio osservare certe ricorrenze, a creare le leggi fisiche?

Penso ad un ipotetico *determinismo economico*. Domani inizio a investire nel *forex*: scambio valute, perdo, guadagno, studio l'andamento dei mercati, adatto la mia strategia... Un economista mi osserva. Se non riesce a individuare alcun pattern nel mio comportamento, concluderà che è *come se* stessi agendo a caso. Se invece individua dei pattern (e li individuerà, perché ho i miei criteri, il mio stile d'investimento), potrà produrre una formalizzazione matematica del mio agire, una formula, un algoritmo che permette di prevedere le mie scelte future nei vari contesti. In questo caso, il mio agire *appare* determinato. Ma, mi chiedo: *sono* davvero determinato? Esiste cioè una qualche legge metafisica che mi impone di comportarmi così come mi comporto? Sarebbe assurdo: posso anche credere all'esistenza concreta di leggi fisiche, ma che esistano anche leggi economiche, in un loro piano autonomo, è un po' troppo lontano dal senso comune. Sono io che, agendo, creo la legge, ma non ne sono vincolato, in futuro potrei modificare le mie strategie d'investimento e così cambiare le leggi economiche che rappresentano il mio agire.

Penso ora ad un *determinismo statistico*. Ogni anno c'è un certo numero di omicidi; posso agire per farlo calare, ma non sono così assurdamente ottimista da credere di poterlo ridurre a zero, è impossibile che per la fine dell'anno proprio nessuno abbia compiuto un omicidio: dovrei quindi dire che qualcuno, anche se non posso sapere chi, è determinato ad uccidere? Ovviamente no. Se lancio mille monete, so che otterrò circa 500 teste e 500 croci: dunque ci sono 500 monete determinate a far uscire testa? No: ogni moneta, presa singolarmente, è davvero indeterminata, mostrerà una faccia a caso, ma in virtù delle sue caratteristiche posso comunque formulare una legge statistica che descrive il comportamento d'insieme di un gruppo di monete; e poiché il comportamento del gruppo emerge dal comportamento delle sue parti, anche tale legge deve essere apparente.

Dunque:

- per ipotesi, almeno in alcuni casi (come nei miei investimenti), io sono libero, ma *appaio* determinato perché creo leggi apparenti (seguissi invece leggi metafisiche, *sarei* determinato);
- per il teorema del libero arbitrio di Kochen, se io sono libero, anche le particelle che mi compongono lo sono;
- essendo io libero, anche le leggi fisiche seguite dalle mie particelle devono essere apparenti, create dal comportamento e senza potere di determinazione.

La maggior parte dei filosofi considera la fisica l'unica scienza *non-speciale*, la scienza alla quale tutte le altre possono essere ridotte: così, accettano che tutte le leggi siano apparenti, tutte *tranne* quelle della fisica, che devono essere concrete e reali in quanto da esse emerge tutto il resto. Ma se io sono libero nel mio agire, se le leggi che seguono sono apparenti, allora anche le leggi fisiche devono esserlo. Questo ha il vantaggio di creare una metafisica più semplice: infatti, dove sono queste leggi metafisiche? Di cosa sono fatte? Ma se le leggi non hanno esistenza concreta, queste domande assurde svaniscono. Il mondo è come un dado che mostra sempre la stessa faccia pur non esistendo alcun fatto che determini quale faccia debba uscire.

Al che devo chiedermi: posso capire se davvero ogni legge è apparente, e non metafisica? C'è un qualche modo di verificare l'argomentazione esposta?

La *teoria del chaos* è un settore matematico in rapido sviluppo. Esistono sistemi, detti *caotici*, che, a piccole variazioni delle loro condizioni iniziali, rispondono con comportamenti molto diversi: se colpisco la sfera da biliardo un millimetro a destra rispetto al suo centro la faccio rotolare verso sinistra, e se la colpisco un millimetro a sinistra la faccio rotolare verso destra, a una piccola differenza nel colpo corrispondono traiettorie molto diverse; e se accendo due fiammiferi, per quanto io possa stare attento a fare tutto nello stesso modo, non otterrò mai due fiamme identiche, che si muovono nello stesso modo, che creano le stesse figure. Il caso più esemplare di sistema caotico è quello del *doppio pendolo*, due pendoli connessi all'estremità: posso sforzarmi di lasciarlo cadere sempre nello stesso modo, eppure ogni volta ruoterà come impazzito disegnando traiettorie diverse e complessissime.

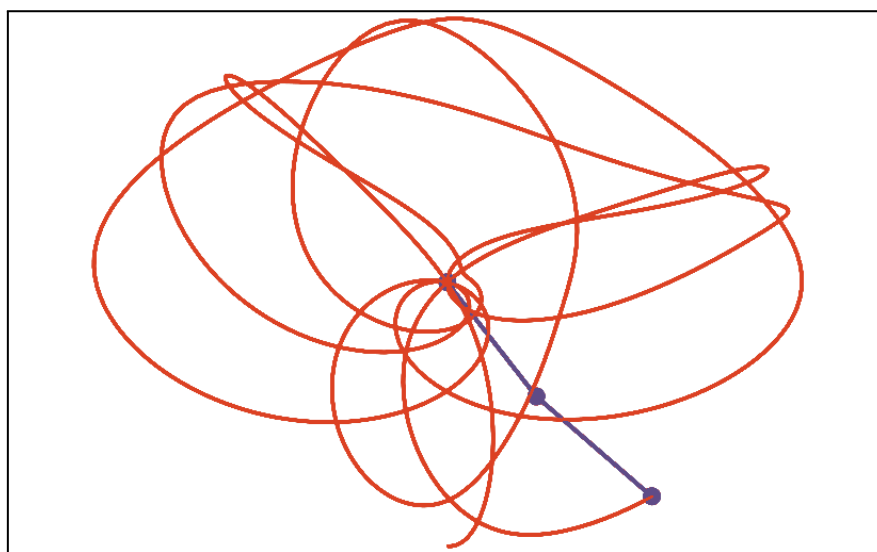


Figura 16

I sistemi caotici sono in realtà determinati: avessi conoscenza esatta delle condizioni di partenza, potrei anche prevedere il loro sviluppo futuro. Invero, qualsiasi sistema, trascorso un certo periodo (detto *tempo di Ljapunov*), diventa caotico: per il sistema solare questo periodo è di cinque milioni di anni, per la rotazione di Iperione (una luna di Saturno) è di 36 anni, per l'attività dei miei neuroni sarà di qualche frazione di secondo. Il *tempo Ljapunov* varia con la precisione delle misurazioni: se misuro la posizione di Mercurio compiendo un errore di anche solo trenta millimetri, la sua orbita diventa imprevedibile dopo 200 milioni di anni, se voglio continuare a fare previsioni affidabili dovrò rimisurare la posizione. Questo accade perché, essendo impossibile tenere in considerazione ogni più piccola influenza, le leggi che i fisici sono in grado di individuare sono approssimazioni, e gli strumenti di misurazione che producono sono imperfetti; ma, se posso prevedere correttamente il comportamento di un sistema per milioni di anni, è legittimo credere che, conoscessi tutte le variabili, potrei prevederlo infinitamente.

Epperò, il fatto che questi sistemi seguano una legge non vuol dire che io sia in grado di individuarla: alcuni sistemi potrebbero essere così complessi da risultare impossibili da comprendere per una mente con capacità di calcolo limitate. In tal caso, il sistema *apparirà* casuale, anche se, avessi adeguate capacità di calcolo e conoscenza delle condizioni iniziali, mi *apparirebbe* determinato. Di conseguenza, dal fatto che io non riesco ad individuare con alcun mezzo pattern di comportamento, non posso concludere che tali pattern non esistano; e quindi non posso neanche essere sicuro che ciò che mi appare casuale lo sia davvero.

Anche in fisica esistono sia interpretazioni della meccanica quantistica con elementi di genuina casualità, sia interpretazioni (come quella del fisico-mistico Bohm) che invece sono del tutto deterministiche. Riuscirò un giorno a scoprire qual è l'interpretazione corretta? Non è detto. Ora come ora, interpretazioni deterministiche e indeterministiche sono empiricamente equivalenti, cioè fanno esattamente le stesse previsioni *testabili*, prevedono gli stessi fenomeni *osservabili*. Forse un giorno, calcolando nuove previsioni e sviluppando strumenti per osservare più fenomeni, scoprirò che un'interpretazione prevede correttamente un certo fenomeno, mentre tutte le altre sbagliano; anche in tal caso, però, potrò emendare e perfezionare le altre interpretazioni per portarle a fare la stessa previsione. Alla fine, potrei arrivare ad avere una teoria deterministica e una indeterministica matematicamente equivalenti, che fanno le stesse previsioni in ogni circostanza, e tali da essere impossibile sviluppare alcuna altra teoria migliore. In tal caso, il mondo resterebbe *o* determinato *o* indeterminato, ma essendo le mie teorie empiricamente indistinguibili, io non potrei capire quale sia la natura del mondo.

Ora, posto un qualsiasi numero di osservazioni, ci sono infinite teorie possibile che, tramite percorsi diversi, spiegano gli stessi fatti; dunque, fino a prova contraria, è molto probabile che io possa spiegare i risultati dei vari test sia tramite una teoria deterministica, sia tramite una teoria con elementi casuali. Se una teoria è deterministica, allora o è composta da leggi apparenti, o rappresenta leggi metafisiche; se invece ci sono fenomeni casuali, non c'è alcuna legge a guidarli. Ma poiché io non posso sapere quale delle due teorie sia vera, non posso neanche sapere se dietro quei fenomeni vi siano leggi oppure no. Al che, seguendo Putnam, in assenza di altre argomentazioni, è legittimo interpretare il mondo sia come guidato da leggi metafisiche (e quindi senza libero arbitrio), sia come privo di leggi metafisiche (e quindi che può prevedere libero arbitrio); ma io ho argomentazioni a favore di una visione del mondo con libero arbitrio, quindi quest'ultima teoria sarà da favorire.

blabla bla

blablabla

blablabla

Il tema del libero arbitrio è sempre stato presente nella riflessione filosofica: tra i classici si possono ricordare il solito "Ricerca sull'Intelletto Umano" di Hume, "Della Libertà e Necessità" di Hobbes, l'"Etica" di Spinoza, "Saggi sulle Facoltà Attive della Mente Umana" di Reid, "Sulla Libertà del Volere" di Schopenhauer (la sua citazione famosa, però, viene da "Come Io Vedo il Mondo" di Einstein, suo estimatore).

L'esperimento di Libet, discusso in suoi paper abbastanza tecnici quali "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action" e "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)", ha avuto un'influenza senza pari nella discussione filosofica del determinismo. Ad esso si rifà spesso Dennett, che in "L'Evoluzione della Libertà" ed "Elbow Room: The Varieties of Free Will Worth Wanting" difende un compatibilismo vicino al determinismo. Altri difensori del determinismo sono Pereboom ("Living Without Free Will", "Free Will, Agency and Meaning in Life"), il sociologo Wegner ("The Illusion of Conscious Will") e Ted Honderich ("On Determinism and Freedom", "How Free are You?", "A Theory of Determinismo").

L'argomentazione della battaglia navale si trova nell'Organon di Aristotele; l'argomentazione del Demone di Laplace si trova nel suo "Saggio Filosofico sulle Probabilità", ma anche nel "Theoria Philosophiae Naturalis" di Boscovich. Prigogine usa la termodinamica moderna per rifiutare il demone di Laplace in "Il Futuro è già Determinato?" (ma si veda anche "La Fine delle Certezze" e "Dall'Essere al Divenire"). Il Principio di Ragion Sufficiente è esposto sia nei "Saggi di Teodicea" che nella "Monadologia" di Leibniz, ma non lo tratta mai in modo sistematico.

Riguardo gli altri tipi di determinismo, i migliori sul determinismo geografico sono "Armi, Acciaio e Malattie" e "Collasso. Come le Società Scelgono di Morire o Vivere" di Jared Diamond. Idee molto diverse, ma sempre deterministe, si trovano nel simile "Why Nations Fail" di Acemoglu e Robinson, economisti eterodossi. Anche certi storici sposano visioni deterministe: anche se Murray non si definisce tale, il metodo storiometrico che sta alla base di "Human Accomplishment" presuppone un qualche tipo di determinismo; similmente, i suoi "The Bell Curve" e "Human Diversity" sono esempi, effettivamente illuminanti, di determinismo genetico e psicologico. Sulla stessa linea si pone "Why the West Rules" di Ian Morris. Precedenti storici di questi testi, oltre che nell'opera di Galton, Terman, Spearman, si possono trovare in "A Statistical Study of Eminent Men" di Cattell, "High IQ Correlates with High Eminence" di Walberg e "Sociobiology" di W. O. Wilson, ma si può risalire anche fino a Wilhelm von Humboldt con libri come "Considerazioni sulle Cause Efficienti della Storia Universale" e "Sulla Differenza della Struttura Linguistica dell'Uomo e sulla sua Influenza sullo Sviluppo Spirituale del Genere Umano". Per un'opinione di senso opposto, si può vedere "Intelligenza e Pregiudizio" di Stephen Jay Gould, e magari "Yet More on the Heritability and Malleability of IQ" del statistico Cosma Shalizi, pubblicato sul suo blog Three-Toed Sloth. Per il determinismo linguistico, può essere curioso "Non Dormire, ci Sono Serpenti" di Everett e "Metafora e Vita Quotidiana" di Lakoff. Infine, per il paradosso della prefazione che ho usato in un'argomentazione, si veda "Epistemology and Inference" di Kyburg.

Il più importante difensore del compatibilismo è Harry Frankfurt: il suo paper "Alternate Possibilities and Moral Responsibility", affrontando il libero arbitrio da un punto di vista etico, tenta di dimostrare che vi è piena compatibilità tra la responsabilità morale e l'impossibilità di agire diversamente da come si agisce (inoltre, è sempre lui che ha proposto la versione psicologica, con desideri di desideri, del compatibilismo di Schopenhauer). Altri compatibilisti sono Susan Wolf ("Freedom Within Reason") e Frithjof Bergmann ("On Being Free"). Anche Strawson, uno dei filosofi più influenti del secolo scorso, in lavori come "Free Will" e "Freedom and Belief" tende al compatibilismo, almeno nel senso che rifiuta il libertarismo ma non si esprime sul determinismo; similmente, John Martin Fischer ("The Metaphysics of Free Will") accetta la compatibilità del determinismo con la responsabilità morale, ma non si esprime sulla sua compatibilità con la libertà. L'argomentazione dello zigote di Mele, che pure difende il compatibilismo, si trova nel suo paper "Free Will and Luck".

Dall'altro lato, il più fervente difensore del libertarismo è Robert Kane, che ha pubblicato un gran numero di lavori sul libero arbitrio, come "Free Will", "Free Will and Values", "The Significance of Free Will". Un altro difensore di tale tesi è Peter van Inwagen ("Thinking about Free Will" e il più importante "An Essay on Free Will"), che affronta il problema da una prospettiva teista. La versione di Finch e Warfield dell'argomentazione di van Inwagen si trova in "The MIND Argument and Libertarianism". Nonostante non sia una tesi comune con il compatibilismo, il libertarismo è difeso anche da alcuni tra i più influenti filosofi del secolo scorso, come Chisholm in "Agents, Causes and Events" e Nozick in "Spiegazioni Filosofiche". L'argomentazione di Einstein sull'esistenza di variabili nascoste, cioè il paradosso EPR, è esposto la prima volta in "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?". La risposta di Bell è "On The Einstein Podolsky Rosen Paradox", scritto a ben 29 anni di distanza, che contiene il suo famoso teorema. Il teorema di Kochen-Specker è esposto in "The Problem of Hidden Variables in Quantum Mechanics". Da questo Kochen, assieme a Conway, deriva il "teorema del libero arbitrio" al quale ho solo accennato brevemente; chi è interessato alla sua dimostrazione, può controllare il paper "The Strong Free Will Problem", che è il più semplice tra i loro paper (in maniera più completa è esposto in "The Free Will Theorem"; il solo Kochen lo ha poi rielaborato in "Born's Rule, EPR and the Free Will Theorem").

L'argomentazione per ragione, la più importante difesa del libertarismo a suo modo già implicita nel teorema di Bell, trova la sua base in "Miracoli - Uno Studio Preliminare" di C. S. Lewis. Argomentazioni simili si trovano anche in "God and the Argument from Mind" e "Scaling the Secular City" di J. P. Moreland, in "Metaphysics" di Richard Taylor, in "Guide to Modern Philosophy" di C. E. M. Joad, in "The Emergent Self" di Hasker, e in "Reasons to Believe" di Richard Purtill, mentre Lewis stesso dice di aver trovato argomenti simili in opere precedenti la sua come "The Foundations of Belief" di Balfour, "Orthodoxy" e "Thomas Aquinas: The Dumb Ox" di G. K. Chesterton, e "Possible Worlds" di J. B. S. Haldane; Victor Reppert ha scritto l'eccellente "C. S. Lewis's Dangerous Idea", in assoluto il libro più approfondito su questa argomentazione. Una critica all'argomentazione di Lewis è esposta da Anscombe in "Metaphysics and the Philosophy of Mind". Plantinga espone la sua versione dell'argomentazione più volte: la prima volta in "Warrant and Proper Function", la seconda in "Warranted Christian Belief", la terza in "Knowledge of God" e la quarta in "Where the Conflict Really Lies: Scienze, Religion, and Naturalism". Un'antologia di forti critiche alla sua versione è "Naturalism Defeated?", il quale contiene un'appendice con ulteriori specificazioni di Plantinga. Tutte queste versioni dell'argomentazione sono, in effetti, poco convincenti e, escluso il caso di Haldane, motivate da interessi religiosi; credo che la versione migliore sia invece quella di Popper, che espone in "The Self and Its Brain" e "Objective Knowledge".

(citazione Stephen

Cave: <https://www.theatlantic.com/magazine/archive/2016/06/theres-no-such-thing-as-free-will/480750/>)

Metafisica: filosofia della matematica

Settimo ragionamento: le metafisiche classiche sono indesiderabili

Questo capitolo è il più tecnico del libro. Come in quello sulla teoria della verità, mi limiterò ad esporre le argomentazioni più importanti, questa volta non perché ci sia già una teoria ampiamente e convincentemente sostenuta, ma perché nessuna proposta risulta soddisfacente. Lo schema quindi è molto semplice:

- P1** le prove a favore del platonismo non sono cogenti;
- P2** i teoremi di Gödel rendono l'aristotelismo improbabile;
- P3** il kantismo richiede di abbandonare la logica classica, il che è indesiderabile;
- C** quindi bisogna favorire teorie alternative.

Le tre tesi: dalla preistoria in avanti

Da un lato la matematica, o meglio la *cognizione* matematica, predata l'uomo: anche gli animali riconoscono insiemi di oggetti e si accorgono quando il loro numero cambia, la mamma-piccione per prendersi cura delle sue uova deve avere una qualche coscienza di quante ne ha deposte e quante ne mancano alla schiusa. Ma dall'altro lato la matematica, o meglio il *ragionamento* matematico, non è universale tra gli uomini: tutt'oggi esistono popolazioni di cacciatori-raccoglitori che non dispongono di parole per indicare quantità superiori ad un certo piccolo numero.

Lo sviluppo dell'agricoltura è di stimolo al ragionamento matematico: infatti le civiltà agricole necessitano della matematica per quantificare il raccolto, gestire le transazioni economiche, calcolare il valore di una certa quantità di grano, perfino le prime approssimazioni di π rispondevano alla necessità di calcolare a fini fiscali l'area di terreni circolari. Inoltre, quelle agricole sono civiltà che guardano in alto, la terra punta il dito verso il cielo: la vita dei campi segue il ritmo degli astri, con le fasi lunari, gli equinozi, i solstizi. Così l'agricoltura va di pari passo con l'astronomia: la matematica è subito applicata per prevedere i moti celesti e tenere conto dei giorni, delle stagioni, degli anni.

Già duemila anni prima di Cristo, egiziani e babilonesi avevano sviluppato un'eccellente conoscenza dei cieli. Proprio l'astronomia di queste civiltà costituisce la radice di tutta la filosofia, il curriculum di studi dei primissimi filosofi: anche Talete aveva imparato in Egitto ad usare la matematica per prevedere le eclissi, la data di inizio delle stagioni e, pare, anche il meteo, in modo

da dedurre la qualità dei prossimi raccolti e così arricchirsi scommettendo sull'andamento del mercato agricolo.

All'epoca di Talete il più importante tema filosofico era la natura dell'*archè*, quella sostanza primigenia che compone tutte le altre e permette loro di esistere. Secondo la visione del mondo classica, l'universo è composto da quattro elementi: acqua, fuoco, terra e aria. Per Talete, che deve aver visto molte volte il mare d'inverno, questo mare che con la sua ruvida calma ispira saggezza, l'*archè* è l'acqua: infatti che il mondo primordiale fosse un caotico oceano è un'idea comune nelle religioni del tempo, anche la Bibbia afferma che prima della creazione «lo spirito di Dio aleggiava sulle acque».

L'allievo prediletto di Talete, Anassimandro, cambia le carte in tavola: poiché acqua e fuoco sono contrari, così come terra e aria, e da una certa sostanza non posso creare il suo opposto, nessuno dei quattro elementi può da solo formare tutti gli altri tre, l'*archè* deve essere invece l'*apeiron*. *Apeiron* significa *indistinto*: l'*archè* è coincidenza degli opposti, un miscuglio indifferenziato di tutte le sostanze che io individuo come sue parti. Epperò *apeiron* significa anche *infinito*, *illimitato*: l'*archè* genera infinitamente infiniti mondi, e quindi ogni cosa ha in sé il principio di tutte le altre. Anassimandro pensa ad una sostanza naturale ma con le caratteristiche tradizionalmente assegnate alle divinità: tenta così di accordare l'*essere* col *divenire*, l'illimitato col limite necessario all'equilibrio. Tutt'oggi le intuizioni di questo antico filosofo, che aveva proposto anche una specie di teoria evolucionistica, impressionano per la loro modernità.

Probabile allievo di Anassimandro è Pitagora, l'uomo che secondo Russell più ha influenzato la Storia del pensiero. Cosa è Pitagora? Ogni definizione pare riduttiva ma, con buona approssimazione, lo definirò un cultista. Aveva formato uno stretto circolo di discepoli che si riuniva in luoghi appartati e al quale era vietato avere rapporti con il resto della società (chi voleva mantenere rapporti con l'esterno poteva ascoltare Pitagora solo da lontano e senza vederlo, nascosto da una tenda). Questi pitagorici mettevano in comune tutti i loro beni, facevano voto di castità, e si dedicavano anima e corpo allo studio di dottrine che giuravano di mantenere segrete: se un discepolo rivelava all'esterno le dottrine pitagoriche, veniva considerato morto, tanto che gli si faceva il funerale nonostante il suo corpo sembrasse ancora vivo. Importante per i pitagorici era anche l'aderenza ascetica a certi precetti stravaganti: non potevano mangiare fave né interiora o certi tipi di pesce e carne, non potevano portare vestiti di lana né anelli, non potevano spezzare il pane, mettersi il sandalo sinistro prima del destro, lavarsi il piede destro prima del sinistro, discutere dottrine al buio, specchiarsi con una fiaccola in mano... La loro vita era dedicata anzitutto allo studio della matematica e delle sue connessioni con l'astronomia e la musica: i pianeti si muovono seguendo ritmi matematici, le note prodotte pizzicando una corda variano secondo rapporti di quantità, e i pianeti stessi, muovendosi, producono una melodia, la *musica delle sfere*. Da ciò, Pitagora implicava anche una serie di dottrine simil-religiose: aveva appreso in Egitto della *metempsirosi*, ed era convinto che, dopo la sua morte, la sua anima sarebbe trasmigrata in un altro corpo. Infine, nonostante le molte restrizioni, i pitagorici ambivano anche al potere politico, e avevano ricoperto cariche governative in varie città dell'Italia meridionale. È un po' come se oggi un qualche monaco formasse un nuovo ordine il cui carisma religioso sta nello studio delle scienze e nell'impegno in politica.

L'*archè* è l'*apeiron*, l'*infinito*; e cos'è che è infinito? I numeri. In un certo senso, per Pitagora il numero è l'*archè*, la quantità sta alla base di tutto e quindi tutto è analizzabile matematicamente: questo è il grande passo che mancava alla filosofia di Anassimandro. I successi matematici di Pitagora, però, sono modesti: non è lui lo scopritore del teorema che porta il suo nome (è qualcosa di molto più antico), e anzi, ironicamente, proprio la matematica ha messo in crisi tutte le sue dottrine. Infatti,

se tutto è un numero, ogni cosa, ogni rapporto, ogni relazione deve essere rappresentabile da un certo numero intero o frazione; esistono però quantità *incommensurabili*, quantità inesprimibili tramite numeri interi o frazioni (per esempio, la diagonale di un quadrato è $lato * \sqrt{2}$, ma $\sqrt{2}$ non equivale ad alcuna frazione), quindi la premessa deve essere falsa. Ciò nonostante, l'intuizione di Pitagora resta fondamentale, perché davvero il moto dei pianeti è rappresentabile matematicamente, e così il rapporto tra le note o tra le forme di un'immagine, e perfino il modo in cui si formano le montagne e i cristalli nelle grotte: aver fede nel fatto che la matematica possa rappresentare correttamente la realtà è la base di tutta la scienza.

L'influsso di Pitagora scorre potente nelle vene di Platone. Inizialmente, però, Platone non voleva essere filosofo, era invece un poeta e uno sportivo: il suo stesso nome si riferiva, non alla larghezza del suo pensiero, ma a quella del suo torace, così esaltata dal suo maestro di lotta libera, sport nel quale eccelleva anche a livello olimpico. Poi ha incontrato Socrate, il pazzo che faceva domande a caso ai passanti, e così si è convinto a diventare filosofo. Anche lui, da uomo saggio qual era, capì subito che senza matematica non è possibile alcuna saggezza: all'ingresso della sua *Accademia* aveva fatto scrivere «non entri chi non conosce la geometria».

La matematica è la più specifica e pura delle materie: del tutto autoreferente, è un sistema di assiomi sui quali i matematici ragionano deducendo conclusioni indipendenti da ogni altro fatto, io potrei fare matematica senza conoscere alcun fatto non-matematico e senza ottenere alcuna nuova conoscenza non-matematica. La filosofia, al contrario, è la materia più generale e impura: il filosofo pesca dati da tutte le altre materie e tutte le commenta. La matematica è puro metodo quantitativo, viene prima di ogni altra materia: i matematici parlano solo a loro stessi, eppure tutti li ascoltano, perché tutte le altre materie hanno bisogno degli strumenti della matematica. Al contrario, la filosofia è puro metodo qualitativo, viene dopo ogni altra materia: i filosofi cercano di parlare a tutti, perché su tutto hanno qualcosa da dire, eppure nessuno li ascolta. Tuttavia, nonostante i filosofi insinuino dubbio in tutto, negando anche la possibilità della conoscenza e la sensatezza dell'etica, nessuna grande scuola filosofica ha mai messo in discussione il valore della matematica. Invero, non c'è alcuna divisione netta tra questi due estremi. Così, quando gli opposti si incontrano, si crea questa materia, la filosofia della matematica, che viene prima ma anche dopo di ogni altra, che è sia ragionamento sui fondamenti di tutte le scienze sia commento su tutte le loro conclusioni.

Le tre tesi: platonismo, aristotelismo, kantismo

La domanda più pressante per ogni metafisico è se esistano oggetti astratti, o se invece esista solo l'universo fisico: nel primo caso c'è conoscenza anche al di fuori delle scienze naturali, nel secondo no e tutte le altre materie o indagano i fondamenti e i presupposti della scienza naturale oppure sono null'altro che passatempi. Ai tre più grandi filosofi della tradizione occidentale corrispondono altrettante risposte a questo dilemma

Secodo Platone, l'universo materiale così come appare ai sensi è illusorio, e la realtà, con tutto ciò che ha valore e può costituire conoscenza, sta al di là della materia. L'uomo, però, osservando l'imperfetto mondo materiale riesce a intuire anche le *forme* perfette che stanno al di là d'esso: vedendo sgangherate figure simil-triangolari in natura (nelle nuvole, nella corteccia degli alberi, nelle conchiglie...) intuisco il triangolo geometrico perfetto, la *forma* del triangolo. Queste *forme* non sono immaginazioni soggettive, sono fatti, sono entità reali situate nell'*iperuranio*, che è il mondo *più vero*. Da un lato, questo platonismo è uno dei fondamenti del pensiero scientifico: se non

credessi nell'esistenza di un mondo *più vero* nascosto oltre il mondo *fenomenico*, se non credessi nell'esistenza di un complessissimo mondo di cellule e atomi nascosto oltre questa mano apparentemente uniforme, concluderei direttamente che la mia conoscenza si esaurisce in ciò che mi è già evidente. Dall'altro lato, però, il platonismo è anche ciò che c'è di più lontano dal naturalismo scientifico: il mondo reale non è quello empirico, ma quello delle astrazioni, dei valori, delle essenze, delle energie vitali.

Il maggior allievo di Platone, Aristotele, è solo in parte d'accordo col suo maestro. Aristotele conferma l'esistenza reale delle astrazioni, per lui (come per la maggior parte dei filosofi di oggi) le proprietà sono reali: quando dico che le mele sono «rosse» lo faccio riferendomi ad un qualcosa, il *rossore*, che esiste ed è presente contemporaneamente in ogni mela rossa. Ciò che Aristotele invece nega è l'*autonomia* dell'esistenza di queste astrazioni: le proprietà non sono certo riposte in un qualche iperuranio oltre l'universo, bensì esistono *negli oggetti*; cioè, mentre per Platone il *rossore* può esistere a prescindere dall'esistenza degli oggetti rossi, per Aristotele il *rossore* esiste solo finché esistono oggetti rossi.

Il terzo grande filosofo della tradizione occidentale, nato oltre 2100 anni dopo gli altri due, è Kant. Come Platone, Kant è un idealista moderato che crede nell'esistenza di un mondo *più vero*, oggettivo, oltre quello fenomenico, un mondo che chiama *noumeno*; come Aristotele, però, nega che le astrazioni abbiano un'esistenza autonoma in tale mondo. Kant crede che ciò di cui si può avere conoscenza sia il solo mondo *concettualizzato*, il mondo risultante dall'applicazione degli schemi mentali alla realtà: di conseguenza le astrazioni che posso conoscere non esistono né nel *noumeno*, né negli oggetti, bensì è la mente umana che le crea tramite i suoi schemi mentali.

A queste tre generalissime scuole metafisiche – platonismo, aristotelismo, kantismo – corrispondono tre gruppi di teorie della filosofia della matematica: così la riflessione sulla natura della matematica è anche riflessione sulle più importanti visioni del mondo della filosofia occidentale.

Per il platonismo matematico, gli enti descritti dalla matematica sono reali, un numero esiste allo stesso modo di un albero o un atomo.

Per l'aristotelismo matematico, gli enti descritti dalla matematica sono reali ma non hanno esistenza autonoma: i numeri sono come *simboli* che rimandano a qualcos'altro, e quindi esistono solo nella misura in cui simboleggiano qualcosa di non-matematico e reale, il numero «6» esiste solo nella misura in cui simboleggia sestetti di oggetti reali, sei sassolini per esempio. Questa è la posizione più intuitiva: almeno fino all'epoca moderna gli uomini hanno sempre pensato al numero come a una pluralità di oggetti.

Per il kantismo matematico, infine, gli enti descritti dalla matematica non sono reali, è l'essere umano che li crea tramite le proprie intuizioni.

In sintesi: i platonisti sono realisti nei confronti della matematica, i kantiani sono antirealisti, gli aristotelici sono realisti in quanto al valore di verità delle proposizioni matematiche e antirealisti in quanto all'esistenza di enti matematici (riconoscono che la matematica dice qualcosa di vero, ma solo in virtù di fatti non-matematici).

Le prove a favore del platonismo non sono cogenti 1: l'argomentazione d'indispensabilità

L'idea che i numeri siano entità reali suona oggi strana, quasi mistica, sembra numerologia, ghematria, cabbala. Cionondimeno, tutti i fenomeni fisici (ma anche alcuni fenomeni economici, sociologici, psicologici e, tramite la *cliometria*, anche storici) possono essere descritti, previsti, e addirittura scoperti col solo ausilio della matematica: osservo una particella, formalizzo il suo comportamento e poi, con la sola manipolazione delle formule, prevedo nuovi fenomeni, mai osservati prima, ma che davvero si verificheranno. Non a caso, da Platone al continentale Badiou, passando per Cartesio e Leibniz, i più grandi filosofi di ogni epoca sono concordi nell'assegnare alla matematica un ruolo fondamentale nella comprensione della realtà. Ora posso dimostrare che, se assegno alla matematica questa capacità di produrre conoscenza, devo anche assegnare esistenza reale agli enti matematici.

Parto da questa domanda: perché la matematica è così efficace nel descrivere il mondo? Se la matematica è astratta e indipendente da qualsiasi altro fatto, perché esiste questa relazione tra formule matematiche e fenomeni fisici? Com'è possibile che, scoprendo cose nuove su un'equazione, io scopra contemporaneamente cose nuove sul mondo, anche a prescindere dalla mia capacità di osservare i fenomeni descritti?

Il fisico Max Tegmark ha una risposta radicale: non è vero che la matematica è astratta e indipendente dal mondo, è l'universo fisico ad essere solo il riflesso della realtà matematica, la matematica funziona perché è la base della realtà, perché solo le strutture matematiche possibili esistono davvero. Questa posizione, detta *matematicismo*, è la variante più estrema del platonismo: Tegmark stesso ammette che la conclusione di tutti i suoi studi scientifici è riassumibile nella frase «Platone aveva ragione», mentre il suo collega Julian Barbour chiama l'universo astratto e atemporale discendente da questa visione «Platonìa».

Posso dunque sostenere il matematicismo con questa argomentazione:

- P1** *Rasoio di Occam*: tra due teorie concorrenti è da favorire la più semplice (ad esempio: se devo spiegare perché un pianeta si muove in un certo modo, cercherò di produrre una teoria che rende merito del movimento analizzando ciò che già so esistere, e non postulando subito l'esistenza di un pianeta inosservato o inosservabile in più);
- P2** la matematica permette di acquisire conoscenze sulla realtà;
- P3** se la matematica rappresenta enti reali, è normale che permetta di acquisire conoscenze sulla realtà, quindi, per abduzione, gli enti matematici sono reali;
- P4** una teoria che postula l'esistenza dei soli enti matematici (*matematicismo*) è più semplice di una teoria che postula l'esistenza sia di enti matematici che fisici;
- C** quindi, per il rasoio di Occam, il matematicismo è la teoria da favorire.

Sollevo subito una facile obiezione. Rispetto al matematicismo, una teoria che prevede solo enti fisici e nessun ente matematico è ugualmente semplice, più coerente col naturalismo, e comunque capace di spiegare il valore epistemologico della matematica (infatti l'abduzione non è cogente). Al che, io favorirò il matematicismo solo se prima mi si dimostra che non posso negare l'esistenza degli enti matematici. Ebbene, ecco l'*argomentazione d'indispensabilità* di Quine e Putnam, che dimostra proprio quanto richiesto:

- P1** io devo avere *impegno ontologico* nei confronti di tutti e soli gli enti indispensabili alle migliori teorie scientifiche;
- P2** gli enti matematici sono indispensabili alle migliori teorie scientifiche;

C quindi devo avere *impegno ontologico* nei confronti degli enti matematici.

Faccio un esempio. Domenica vado a messa, e sono testimone di un miracolo sconvolgente: il pane e il vino, dal nulla, pur mantenendo in tutto il loro aspetto esteriore, si trasformano di colpo nel corpo e nel sangue di nostro Signore Gesù Cristo. Formulo una teoria per spiegare questo evento: «Dio ha transustanziato il pane e il vino». Al che, posso io credere contemporaneamente a questa teoria e alla proposizione «Dio non esiste»? Ovviamente no: Dio è un ente indispensabile alla mia teoria (perché non potrei crederla senza credere in Dio), quindi devo avere impegno ontologico nei Suoi confronti, devo crederLo realmente esistente. Amen.

Ora, le migliori teorie scientifiche mi dicono che «la forza è il prodotto di massa e accelerazione». Ovviamente devo avere impegno ontologico nei confronti della forza, della massa, e dell'accelerazione: io credo nella forza, anche se non sono uno Jedi. Posso, credendo a questa teoria, negare l'esistenza degli enti matematici? No, perché la forza è un prodotto, e un prodotto è un numero, se non credo ai numeri non posso neanche credere che la forza sia un prodotto; quindi, dovendo credere alle migliori teorie scientifiche, devo anche avere impegno ontologico nei confronti degli enti matematici.

Altro esempio. Il maglione di lana mi molla una scossa. La scienza mi offre la miglior spiegazione per questo fenomeno: «gli elettroni passano dal corpo con carica positiva a quello con carica negativa». Devo dunque avere impegno ontologico nei confronti dell'elettrone, in quanto la sua esistenza è necessaria alla verità della teoria. Ma cos'è l'elettrone? In fisica, è definito da una serie di formule: quindi, se credo «l'elettrone esiste», credo anche l'equivalente «il tal insieme di formule esiste», quindi credo nell'esistenza di enti matematici.

La più forte obiezione all'argomentazione d'indispensabilità si deve all'eroico filosofo Hartry Field: secondo lui la matematica è utile alla chiarezza e alla sintesi ma niente affatto necessaria alla formulazione delle teorie scientifiche; e così, forte di questa convinzione, ha baldanzosamente riformulato una gran fetta della fisica newtoniana senza usare la matematica, dimostrando in modo pratico la sua dispensabilità. In realtà Field usa comunque la geometria, ma la interpreta in modo *nominalistico*: per esempio, usa i punti, ma li interpreta come enti fisici, non come astrazioni; così, la verità delle sue proposizioni geometriche richiede solo l'esistenza di certi enti fisici (poiché tale è il loro riferimento), e non di enti matematici.

Field però interpreta l'argomentazione d'indispensabilità in modo più forte di quanto Quine e Putnam intendessero. Quando affermano che un certo ente è *indispensabile* intendono dire che una volta negata la sua esistenza diventa impossibile produrre una teoria *migliore* di quella attualmente disponibile (*migliore* secondo i criteri della scienza: semplicità, precisione, utilità pratica...). Field dimostra che è possibile riformulare una teoria fisica senza appellarsi ad entità matematiche, ma non dimostra che una teoria di questo tipo possa essere *migliore*.

Anzi, in realtà, Field non dimostra neanche che gli enti matematici siano *sempre* eliminabili: ha sì dimostrato che la fisica newtoniana può essere riformulata, ma secondo molti filosofi ciò non è estendibile alla fisica quantistica (che prevede per esempio l'uso dell'infinito, ed essendo l'esistenza di infiniti enti fisici controintuitiva non pare interpretabile in modo nominalistico). E se anche la fisica quantistica fosse riformulabile, comunque potrebbero esservi altri settori della matematica non ancora applicati alla fisica, ma applicabili in futuro, tali da risultare non interpretabili in modo nominalistico né traducibili in un linguaggio non-matematico.

Ora, nella sezione di meta-etica ho esposto la tesi per la quale io sono giustificato a credere all'esistenza di un certo ente inosservabile solo se è possibile causa di un fenomeno e costituisce la sua miglior spiegazione, cioè solo se può essere addotto: sono giustificato a credere all'esistenza degli elettroni, anche se non posso vederli, perché sono la miglior spiegazione dell'esistenza dei fulmini, che invece posso ben vedere. Quindi io posso essere giustificato a credere all'esistenza degli enti matematici solo se mi è necessario addurli per spiegare un'osservazione. Orbene, ecco un esempio di teoria nella quale gli enti matematici fungono da spiegazione, e perciò risultano indispensabili.

Magiccada è un genere di cicale nordamericane con un ciclo vitale particolarmente lungo: le *Magiccada tredecim*, *neotredecim*, *tredecassini* e *tredecula* si sviluppano nel sottosuolo per 13 anni, le *Magiccada septendecim*, *septendecula* e *cassini* per ben 17 anni, poi emergono in superficie. Perché proprio 13 o 17 anni, e non 15 o qualche altro numero? Salta subito all'occhio che sono entrambi *numeri primi*, numeri divisibili solo per uno e sé stessi. Immagino che oggi diversi insetti, con diversi cicli vitali, depongano le loro uova. La cicala che emergerà tra 13 anni si ritroverà assieme a quali altre specie, tra quelle che hanno deposto le uova oggi? Solo quelle che si presentano tutti gli anni o una volta ogni 13 anni. E se invece avesse avuto un ciclo vitale di 12 anni? Si sarebbe ritrovata con le specie che si presentano ogni 1, 3, 4, 6, 12 anni. Avere un ciclo vitale lungo un numero primo di anni permette quindi di evitare la competizione, poiché pochi altri cicli vitali saranno sincronizzati col proprio.

Dunque, ho osservato un fenomeno, il fatto che le cicale periodiche abbiano un ciclo vitale di 13 o 17 anni. Qual è la spiegazione del tal fenomeno? Il fatto che 13 e 17 sono numeri primi, il loro non avere *fattori non-banali*. E questa è una spiegazione puramente matematica, non fisica o biologica o altro; quindi, in virtù di questo ruolo esplicativo, sono giustificato a credere all'esistenza degli enti matematici.

Ma questa difesa dell'argomentazione d'indispensabilità è tanto originale quanto debole. Gli enti matematici, se esistono, pur essendo non-interagenti col mondo fisico e quindi non-esperibili, possono essere conosciuti in vari modi (quelli già visti in meta-etica: tramite ragione o una qualche facoltà intuitiva); tuttavia, in questo caso, l'abduzione pare insufficiente alla conoscenza. L'economista che ha analizzato il mio agire sui mercati e ha trovato il modo di descriverlo tramite formule *può* porre tali formule come causa e spiegazione del mio agire, ma se, come detto, ciò non implica necessariamente l'esistenza di leggi (in quanto potrei invece essere io a crearle), ancor meno implica l'esistenza di enti matematici. Similmente, che un certo ciclo vitale abbia meno sovrapposizioni con gli altri cicli vitali possibili è un dato empirico che non richiede certo l'esistenza dei numeri, le cicale potrebbero avere un ciclo vitale lungo un numero primo di anni anche se i numeri non esistessero in modo fattuale. E se sostenessi che lo spaziotempo stesso ha natura matematica, che lo scorrere degli anni necessita di per sé degli enti matematici? Sarebbe comunque insufficiente a convincere quei filosofi e fisici che negano l'esistenza reale dello spaziotempo.

Le prove a favore del platonismo non sono cogenti 2: olismo e naturalismo

Quine diceva di preferire le metafisiche desertiche, quelle con pochissimi enti, le metafisiche che riducono i vari concetti a particelle elementari invece di postulare una pluralità di astrazioni, valori, numeri, essenze e quant'altro. E in effetti Quine (come Putnam) non era un platonista in senso pieno. Perché dunque ha sviluppato l'argomentazione d'indispensabilità?

Anzitutto per il suo *olismo*. Quine era un deflazionista e un coerentista: come il senso di una frase è comprensibile solo in relazione ad un intero linguaggio, così le teorie scientifiche sono dimostrate o rifiutate nel loro insieme. Per esempio, la proposizione «l'acqua in pianura bolle a 100 °C» può essere vera o falsa solo in relazione ad una serie di altre proposizioni in merito agli strumenti di misurazione, al loro funzionamento e ad altri fattori, cioè solo in relazione ad un'intera teoria; così, se dimostro la verità di «l'acqua in pianura bolle a 100 °C» automaticamente dimostro anche la verità di tutte le altre proposizioni di quella teoria. Di conseguenza, essendo la matematica onnipresente nel sapere scientifico, è impossibile preservare la verità delle proposizioni scientifiche senza preservare quella delle proposizioni matematiche, se tolgo la pietra d'angolo che è la matematica viene giù tutto il tempio della scienza e devo ricostruirlo con teorie diverse.

Quine, però, è anche uno dei pochi filosofi analitici ad aver preso sul serio il settore della *metafilosofia*, la riflessione filosofica sui metodi e sugli scopi della filosofia stessa. La sua posizione è quella, oggi scontata, del *naturalismo metafilosofico*: il filosofo deve porsi in continuità con la scienza, deve accettare le teorie scientifiche e trarne le sue conclusioni, non può metterle in dubbio o proporre teorie alternative, può al massimo riflettere sui presupposti della scienza e sulla sua metodologia.

Sono effettivamente molti i matematici e i fisici che sposano forme di platonismo: Gödel e Penrose sono due esempi rilevanti in quanto, oltre che scienziati tra i più grandi, sono anche filosofi di alto livello. Dunque, se è vero che gli scienziati credono nell'esistenza degli enti matematici, il filosofo naturalista deve mettersi in continuità e partire anche lui dal presupposto del platonismo.

Ma davvero gli scienziati credono questo, davvero hanno impegno ontologico nei confronti della matematica? Secondo l'epistemologo cristiano Bas van Fraassen non è così.

Van Fraassen segue l'*instrumentalismo*: per lui la scienza è solo uno strumento utile a fare previsioni, non un metodo per conoscere ciò che è inosservabile, la scienza ha la pretesa di prevedere lo sviluppo dei sistemi fisici e non di rendere manifesta la più intima natura della realtà. Questo mi libera dall'impegno ontologico: posso dire, per esempio, che l'elettrone è solo un simbolo utile ad immaginare un meccanismo che produce fulmini come fa la realtà, anche se magari la realtà arriva a produrre fulmini in tutt'altro modo. Nell'*instrumentalismo*, il mondo è come una calcolatrice che mostra il numero «4»: gli scienziati si interrogano sul perché mostri quel numero e tramite un metodo convenzionale producono la spiegazione «ha calcolato 2+2», anche se in realtà potrebbe aver calcolato l'equivalente «1+3»;

È vero che gli scienziati credono di poter scoprire qualcosa sulla realtà, ma ciò non significa che abbiano davvero impegno ontologico verso tutti gli enti delle loro teorie. Anzi, che effettivamente non sia così lo si vede dal fatto che, a volte, pongono presupposti semplificati *che sanno essere falsi*: per esempio, i cosmologi possono simulare l'evoluzione di un insieme di galassie rappresentando le stelle come punti (ignorando quindi le particolarità delle singole stelle, i pianeti, gli asteroidi e altri elementi considerati trascurabili). Affinché le previsioni di questa simulazione siano esatte è necessario che le stelle siano davvero puntiformi; i cosmologi, però, sanno che non è così, né vogliono formulare una teoria che produca previsioni perfette, vogliono invece formulare un'*approssimazione*, la verità della teoria sta nel *comportamento generale* e non nei dettagli. Per esempio, se intendo la teoria «in questa stanza la temperatura è di 25 °C» come approssimata, la rappresenterò meglio tramite una disgiunzione: «la temperatura è di 25 °C, o inferiore a 25,5 °C, o superiore a 24,5°». Ci sono dunque più fatti che possono rendere vera tale disgiunzione, ma poiché non possono darsi tutti insieme, io non ho neanche impegno ontologico verso un qualche

specifico fatto, nessun fatto è di per sé indispensabile. Per lo stesso motivo i cosmologi non credono certo all'esistenza di stelle puntiformi, pur usandole in teorie che ritengono essere vere.

Ora posso dimostrare che gli scienziati hanno un atteggiamento antirealista anche nei confronti degli enti matematici. Ecco una *matrice*, una struttura ordinata di valori:

$$(a \ b \ c \ d \ e \ f \ g \ h \ i)$$

Questa è una matrice *quadrata*, cioè che ha tante righe quante colonne. Se in una matrice quadrata i valori della diagonale principale ($a, e, i...$) sono 1, e tutti gli altri sono 0, ottengo la *matrice identità*:

$$(1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1)$$

Posso anche moltiplicare tra loro le matrici per calcolare un *prodotto matriciale*, così:

$$(a \ b \ c \ d) \times (e \ f \ g \ h) = (ae + bg \ af + bh \ ce + dg \ cf + dh)$$

Se due matrici sono *quadrato* e il loro prodotto matriciale è la *matrice identità*, allora sono *invertibili*: posta una matrice quadrata A , posso *invertirla* trasformandola in quella matrice B che, moltiplicata per la matrice A , produce la *matrice identità*.

Voglio dimostrare che una matrice non-quadrata non è invertibile. Per fare un esempio numerico, l'inverso di 5 è $\frac{1}{5}$ in quanto $5 \times \frac{1}{5} = \frac{1}{5} \times 5 = 1$. Similmente, se io ho una matrice A e B è il suo inverso, allora $(A \times B) = (B \times A) = I$ dove I è la matrice identità. Per ipotesi, A è una matrice non-quadrata, con x righe e y colonne dove $x \neq y$. La matrice che costituisce il prodotto matriciale avrà il numero di righe del primo fattore e il numero di colonne del secondo fattore: per esempio, se A è una matrice 2×4 e B è una matrice 1×3 , il prodotto matriciale $A \times B$ sarà una matrice 2×3 . Dunque, essendo la matrice identità una matrice quadrata, se A ha x righe e y colonne, il suo inverso B avrà y righe e x colonne, perché solo così il loro prodotto avrà un uguale numero di righe e colonne. Epperò: il prodotto matriciale $A \times B$ sarà una matrice con x righe e x colonne, mentre $B \times A$ sarà una matrice con y righe e y colonne. Di conseguenza, $(A \times B) \neq (B \times A)$ in quanto non producono la stessa matrice identità, quindi A e B non sono invertibili.

Orbene, in fisica molti sistemi sono descrivibili da matrici. Immagino di avere in mano un sistema fisico descritto dalla matrice A , uno *slime* che quando viene toccato reagisce vibrando. Io posso prevedere la reazione al tocco, cioè l'evoluzione nel tempo dello slime, eseguendo operazioni sulla matrice che lo descrive: A mi porta dallo slime in stato di quiete allo slime vibrante. Di contro, se mi capita di osservare direttamente lo slime vibrante, vorrò sapere come si è arrivati a questo punto: posso allora riavvolgere la cassetta dell'universo invertendo la matrice, la matrice inversa B mi porterà dallo slime vibrante allo slime in stato di quiete e mi permetterà di capire quale evento passato ha causato lo stato presente del sistema. Ma ecco il problema: in fisica, anche se la matrice che descrive il sistema è non-quadrata, io posso invertirla lo stesso, posso trasformare A in B anche se $A \times B$ e $B \times A$ producono matrici identità di grandezza diversa.

Questo mi stimola due osservazioni. La prima è che matematici e fisici usano una matematica leggermente diversa, i fisici lavorano con assiomi aggiuntivi e si prendono maggiori libertà: ma quindi, se la matematica è reale, quale matematica lo è, quella in cui è inaccettabile invertire una matrice non-quadrata o quella in cui è accettabile? La seconda osservazione è che i fisici sanno benissimo che in algebra lineare una matrice non-quadrata non può essere invertita, epperò sono disposti ad approssimare e a rinunciare ad un po' di rigore per sviluppare un formalismo utile a rappresentare i fenomeni fisici; ma come l'astronomo sarebbe incoerente ad accettare l'esistenza dei punti quando sa che questi sono in realtà stelle, così per un fisico sarebbe incoerente avere

impegno ontologico verso gli enti matematici che sta utilizzando quando sa che, in un certo senso, la sua operazione è matematicamente illegittima.

Il naturalismo metafilosofico, quindi, non supporta il platonismo matematico in quanto gli scienziati stessi non hanno impegno ontologico verso gli enti matematici.

Posso obiettare che, anche se nei casi esposti gli scienziati paiono non avere impegno ontologico nei confronti degli enti matematici, comunque in altre occasioni manifestano comportamenti che suggeriscono il contrario. La grande maggioranza degli scienziati ha una visione del mondo realista, e nel realismo una proposizione è vera se corrisponde ai fatti. Se i matematici non avessero impegno ontologico nei confronti dei numeri, sosterrebbero o una *error theory* per la quale tutte le proposizioni matematiche sono false, o un *quasi-realismo* per il quale i matematici trattano le loro proposizioni *come se esprimessero fatti* anche se in realtà si limitano a portare avanti un'*utile finzione*. Eppure, basta chiedere a un qualsiasi matematico di professione, e non dirà né che $2+2=4$ è una proposizione falsa (nessuno lo direbbe!), né che bisogna *fin gere* di ritenerla vera, confermerà invece di essere un realista che considera le affermazioni matematiche come fattualmente vere; quindi, per essere coerente, *dovrebbe* credere agli enti matematici.

Anche accettando tale obiezione, però, rimane difficile accordare naturalismo e platonismo.

Il naturalismo metafilosofico di Quine, infatti, comprende anche una particolare *epistemologia naturalizzata*. L'epistemologia tradizionale è un cane che si morde la coda: come posso scoprire il metodo ideale per acquisire conoscenza senza avere già un metodo ideale per acquisire conoscenza? Come può qualcosa d'incerto come la filosofia creare una fondazione certa per la ricerca? Ma la scienza è già il miglior metodo d'indagine noto, perciò, se voglio capire com'è che sia possibile sapere qualcosa, non devo interrogare la filosofia, ma la scienza stessa: la scienza non ha bisogno della filosofia per analizzare le sue stesse basi. Di conseguenza le domande epistemologiche, puntando a comprendere come la mente umana arrivi a produrre il sapere scientifico, sono da rimettere alla scienza stessa, nello specifico alla psicologia: il filosofo deve chiedersi non cosa sia la conoscenza, bensì in che modo gli stimoli sensoriali formino le credenze, una domanda questa da indagare empiricamente tramite test psicologici.

Bene, cosa dice la psicologia sui matematici che credono nel platonismo? Dice che per l'essere umano è molto difficile dedicare tutta la propria vita e i propri titanici sforzi ad un qualcosa che si crede non esistere: come per un prete che ha perso la fede è difficile continuare a dir messa, così per un matematico che non crede in un remae metafisico dei numeri è difficile continuare a fare ricerca. Per questo i matematici tendono al platonismo, e non perché la loro opinione sia più informata e razionale.

Sorge un nuovo contrasto. Generalmente l'autorità segue la specializzazione: l'autorità nel determinare se una certa sostanza sia un amminoacido appartiene al chimico, e non al non-chimico. Al che, chi è che ha autorità nel determinare se gli enti matematici esistano o meno: il matematico, che è specializzato nei numeri, o lo psicologo, che mi spiega perché il matematico ha un bias verso il platonismo ed è dunque inaffidabile? La risposta non è banale come nel caso dell'amminoacido. Chi è che ha autorità nel determinare se Dio esista: il teologo, la cui opinione è scontata in quanto se non credesse in Dio non sarebbe diventato teologo, o lo psicologo, che mi dice che il teologo crede per motivi irrazionali e non va seguito? Io devo seguire l'autorità della scienza, ma di quale scienza?

Sembra, in effetti, che non esista un'autorità in merito all'*esistenza* di Dio, non c'è una materia che abbia qualcosa in più delle altre da dire in merito; stessa cosa per l'*esistenza* dei numeri, il matematico non pare avere di per sé più autorità del filosofo su questo argomento, e il naturalismo impone di seguire i matematici solo nei campi nei quali hanno autorità. Nonostante il platonismo nei confronti degli enti astratti sia oggi una posizione molto comune, è relativamente poco sostenuta dai filosofi della scienza, che però sono tra i più propensi a difendere il naturalismo metafilosofico.

Infine, un ultimo problema. L'argomentazione d'indispensabilità sfrutta sia il naturalismo metafilosofico, che impone di credere alle teorie scientifiche, sia l'olismo, che impone di credere a *tutti* gli enti indispensabili a tali teorie (perché, se credo ad una loro proposizione, credo a tutte). Il naturalismo metafilosofico implica l'epistemologia naturalizzata, per la quale la scienza ha autorità ed autonomia nel risolvere le questioni epistemologiche. Se è così, però, naturalismo e olismo entrano in contraddizione: infatti gli scienziati non supportano l'olismo, contrariamente a Quine non credono che il dimostrare la verità di una proposizione dimostri automaticamente anche la verità della teoria d'appartenenza (immagino, per esempio, una teoria che preveda correttamente l'orbita di Plutone e che affermi che la Luna è fatta di formaggio). Così, da un lato, chi accetta il naturalismo dovrebbe rifiutare l'olismo, e con esso anche l'argomentazione d'indispensabilità in quanto diventa legittimo per il filosofo non avere impegno ontologico verso *tutti* gli enti delle teorie; dall'altro, chi accetta l'olismo si pone come non-naturalista, e ugualmente dovrebbe rifiutare l'argomentazione d'indispensabilità in quanto diventa legittimo per il filosofo non credere alle teorie scientifiche.

Tra queste due possibilità, conviene sostenere che il platonismo punti ad una filosofia non-naturalista; l'olismo, infatti, è utile a spiegare come sia possibile conoscere enti matematici non interagenti col mondo fisico: poiché le teorie scientifiche sono giustificate nell'insieme e non proposizione-per-proposizione, le proposizioni matematiche sono giustificate dal loro appartenere ad una teoria giustificata, si giustificano giustificando le altre proposizioni scientifiche (quindi la matematica non è indipendente dalle altre scienze, può essere revisionata se certe scoperte lo necessitano).

La tesi aristotelica 1: cos'è il riduzionismo

Nell'esempio dell'astronomo, i punti sono simboli delle stelle, la teoria approssima il mondo ponendo un punto lì dove in realtà c'è una stella: per questo l'astronomo può sia sostenere l'olismo sia negare l'impegno ontologico nei confronti degli enti geometrici della teoria. Alla luce di ciò, un aristotelismo matematico, una filosofia nella quale gli enti matematici esistono ma solo come simboli di altri enti non-matematici, pare una posizione moderata e di buon senso.

Una filosofia aristotelica richiede di *tradurre* questa collezione di simboli che è la matematica in un altro linguaggio: come il simbolo del semaforo rosso è equivalente alla proposizione «devi fermarti», così ogni proposizione matematica deve risultare equivalente ad un'altra proposizione i cui riferimenti sono gli enti non-matematici simboleggiati. In questa classe di teorie, come il simbolo del semaforo rosso è solo un modo tra i tanti possibili per comunicare un certo messaggio, così la matematica è solo uno strumento che rappresenta qualcosa di rappresentabile (sia pur in modo meno chiaro e veloce) anche tramite altri mezzi.

La prima teoria aristotelica è quella del *logicismo*, sviluppata da Frege e Russell. Il suo scopo è ridurre la matematica alla logica, dimostrare che ogni enunciato matematico è riscrivibile utilizzando solo gli strumenti della logica. Non sarà un progetto insufficiente, troppo modesto? In fondo, non è forse vero che la logica, essendo essa stessa un settore della matematica, andrebbe poi a sua volta ridotta, riscritta in altri termini? No, non è proprio così.

I settori della matematica sono innumerevoli, e studiano di tutto:

- l'*algebra* studia i simboli matematici e le loro regole d'utilizzo;
- la *teoria dei numeri* studia incredibilmente proprio ciò che dice il suo nome (ma in particolare i primi)...
 - ...e include anche l'*aritmetica*, che studia le operazioni fondamentali;
- la *geometria* studia le forme tracciabili nello spazio...
 - ...e da essa discende la *topologia*, che studia quelle proprietà che gli oggetti geometrici conservano anche quando deformati (per esempio, un cerchio trasformato in un quadrato rimane sempre una figura chiusa);
- l'*analisi* usa integrali, derivate, limiti, equazioni differenziali e altri strumenti per scomporre gli oggetti in infinitesimi e così studiare varie proprietà, come il cambiamento nel tempo;
- i vari campi della *matematica applicata*, infine, studiano problemi pratici di interesse anche per le altre materie:
 - la *meccanica fisica* studia il comportamento di oggetti solidi, fluidi, particelle, infrastrutture e quant'altro;
 - la *statistica* studia la probabilità e le strutture di insiemi di dati;
 - la *crittografia* studia le tecniche di comunicazione sicura, come far sì che un messaggio giunga ai destinatari desiderati e a nessun altro;
 - la *combinatoria* studia certe strutture, soprattutto pattern, in insiemi di oggetti;
 - la *teoria della decisione* modella vari contesti possibili secondo i parametri d'interesse per individuare quello di valore massimo;
 - l'*informatica* mi permette di scrivere ciò che sto scrivendo.

Dov'è la logica? La logica sta, con l'*insiemistica*, nel settore dei *fondamenti della matematica*. Come nota Russell, sin dall'antica Grecia la logica è sempre stata considerata una materia a parte: è vero che, in quanto studio del ragionamento corretto, la logica è necessaria alla matematica e manifesta nella sua pratica; ma per i greci, ancor prima di questo, la logica è una materia inerente al linguaggio. Quando espone le sue teorie logiche, Aristotele parte dal linguaggio naturale, parte da un teorema specifico come «Socrate è un uomo, se è un uomo è mortale, quindi è mortale», poi astrae, generalizza sostituendo i nomi con variabili: « x , se x allora y , quindi y ». Come conseguenza di questo metodo, i principi della logica saranno accettati o rifiutati a seconda del modo in cui uno parla, si sosterranno diverse logiche a seconda di ciò che *si percepisce come asseribile*. Per questo, in epoca moderna, logica e matematica si avvicinano sempre più: come la matematica parte da astrazioni che solo successivamente applica ai casi concreti, così la logica parte da assiomi formali che solo successivamente cerca di tradurre in lingua naturale.

Da un lato, è corretto dire che la logica è matematica: la tesi del logicismo è appunto che vi sia equivalenza tra le due. Dall'altro, però, non si deve intendere il progetto logicista come la riduzione della matematica ad un suo stesso settore: la logica è una scienza formale a parte, come la *storia della matematica* è solo storia (che discute di matematica), così la *logica matematica* è solo logica (che rappresenta la matematica). Quale sia poi la natura della logica, è una questione che sta al di fuori di questo settore della filosofia: i logicisti sono interessati solo alla natura della matematica, la loro

posizione sulla natura della logica (se sia riducibile a processi mentali, alla razionalità, ad algoritmi neurologici) è puro agnosticismo.

Ora voglio cercare di capire l'aspetto della teoria che il logicismo cerca di produrre, voglio vedere che tipo di affermazioni dovrebbe contenere il libro intitolato «la matematica ridotta alla logica».

Ogni sistema matematico, per esistere, necessita di almeno tre elementi:

- dei concetti primitivi (non definiti da altri concetti);
- una logica, cioè delle regole che spieghino come ricavare conclusioni da premesse;
- e, appunto, delle premesse, dei *postulati* autoevidenti, composti usando i concetti primitivi.

Come concetti primitivi da porre alla base di tutta la loro matematica, gli antichi greci avevano scelto quelli geometrici di *retta* e *punto* (più una manciata di relazioni come *stare su*, e poco altro). Basandosi sulla logica classica dell'epoca di Aristotele, Euclide era riuscito a ridurre tutta la geometria a cinque postulati:

1. posti due punti qualsiasi, tutte le rette che posso tracciare tra quei due punti sono sovrapposte (cioè posso tracciarne solo una);
2. un segmento può essere prolungato indefinitamente;
3. dato un punto e una lunghezza (cioè un centro e un raggio), posso descrivere un cerchio;
4. tutti gli angoli retti sono sovrapponibili;
5. posto un punto esterno ad una retta data, tutte le rette parallele a questa passanti per quel punto sono sovrapposte (cioè c'è una sola parallela).

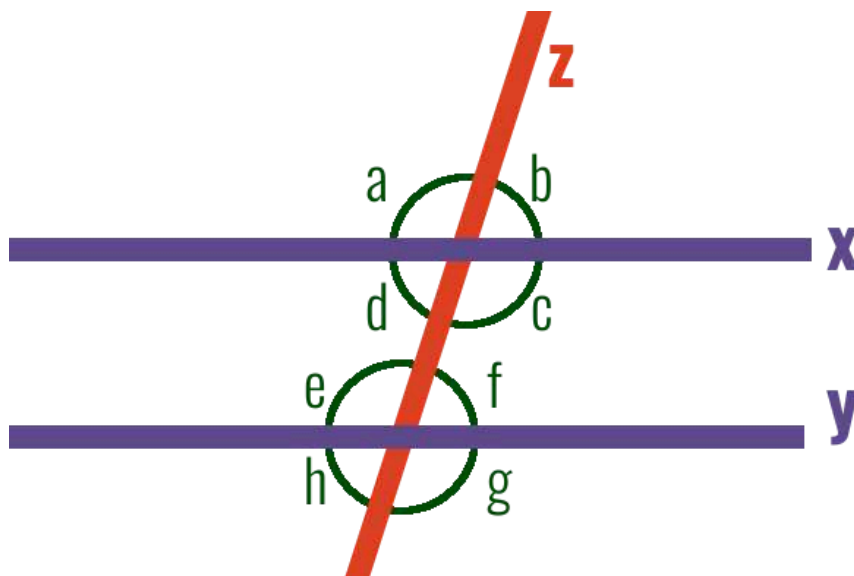
Per semplicità, ho adottato una versione divulgativa del quinto postulato e ho usato direttamente alcuni concetti non-primitivi, i quali però sono sempre definiti dai soli primitivi.

Definizioni.

- un *segmento* AB è l'insieme di punti di una retta compresi tra i punti A e B ;
- la *lunghezza* di un segmento è la classe di tutti i segmenti sovrapponibili a quello di riferimento;
- l'*angolo* è ognuna delle parti in cui lo spazio è diviso da due segmenti aventi un punto in comune (dove lo *spazio* è l'insieme di tutti i punti);
- l'*angolo retto* è la metà di un *angolo piatto*, che è l'angolo formato da due segmenti tali da essere uno il prolungamento dell'altro (cioè i cui punti appartengono alla stessa retta);
- il *cerchio* è l'insieme di tutti i punti aventi una determinata distanza da un certo altro punto (dove la distanza tra due punti A e B è la lunghezza del segmento AB);
- due rette sono *parallele* se non hanno punti in comune o se sono sovrapposte (se hanno un solo punto in comune sono *incidenti*).

Queste definizioni non costituiscono nozioni aggiuntive ai cinque postulati, bensì si limitano a *battezzare* una certa configurazione di rette e punti in una parte di spazio.

Dunque, se un essere eterno conoscesse solo questi cinque postulati e la logica aristotelica, comunque prima o poi riuscirebbe a dedurre ogni singola proposizione geometrica, chi conosce questi cinque postulati conosce implicitamente tutta la geometria. E ora farò un esempio per mostrare come ciò sia possibile.



Teorema. Poste due rette parallele x e y , se la retta z è parallela a x , allora è anche parallela a y ; se invece z è incidente con x , allora è incidente anche con y .

Dimostrazione.

Primo caso: x è parallela a y e a z . È possibile che z sia incidente con y ? Ipotizzo lo sia, e chiamo p il punto d'incontro. Di conseguenza per il punto p passano due rette, y e z , entrambe parallele a x . Per il quinto postulato, ciò è possibile solo se sono sovrapposte, ma se sono sovrapposte sono parallele anziché incidenti. L'ipotesi porta ad una contraddizione e va quindi rifiutata: z deve essere parallela a y .

Secondo caso: x è parallela a y , z è incidente con x . È possibile che z sia parallela a y ? Ipotizzo lo sia. Esiste un solo punto p in cui z interseca x , e per quel punto passano due rette, z e x , entrambe parallele a y . Di nuovo, per il quinto postulato ciò è possibile solo se sono sovrapposte, ma se sono sovrapposte non sono incidenti. L'ipotesi va quindi rifiutata: z deve essere incidente con y .

QED

Conoscendo solo i postulati di Euclide e i principi della logica, battezzate *parallele* e *incidenti* le uniche due possibili configurazioni di rette (perché o si incrociano o non si incrociano), sono arrivato a dedurre questo teorema.

Più di 2100 anni dopo la nascita di Euclide, un grande matematico e glottoteta italiano, Giuseppe Peano, riesce nell'impresa di ridurre tutta l'aritmetica dei numeri naturali a cinque assiomi:

1. lo 0 è un elemento dell'insieme dei numeri naturali;
2. esiste una *funzione successore* S che accoppia i numeri in modo tale che, posto un certo numero naturale, anche il suo successore è un numero naturale;
3. la *funzione successore* S è *iniettiva*, cioè posti due numeri naturali qualsiasi x e y diversi tra loro, il successore di x non è anche il successore di y ;
4. per ogni numero naturale, il suo successore è diverso da 0;
5. *principio d'induzione*: se un qualsiasi insieme di numeri naturali include 0 e il successore di ogni suo elemento, allora quell'insieme include *tutti* i numeri naturali, cioè l'insieme dei numeri naturali è il più piccolo insieme che contenga 0 e il successore di ogni suo

elemento, perciò se 0 e i suoi successori hanno una certa proprietà allora *tutti* i numeri naturali hanno quella proprietà.

Gli assiomi 3 e 4 impediscono alla funzione S di formare circoli; questo, in combinazione con l'assioma 2, implica l'infinità dei numeri naturali, che per l'assioma 5 sono solo 0 e i suoi successori. La funzione S , inoltre, permette di definire i simboli normalmente utilizzati in matematica: «1» è il successore di 0, cioè $S0$, «2» è il successore di 1, cioè $SS0$, eccetera.

Faccio ora un esempio relativo alle operazioni fondamentali.

Definizioni. Si dice *somma* la funzione tale che, posti due numeri qualsiasi m ed n :

1. $m+0=m$
2. $m+Sn=S(m+n)$

Teorema. Assurdamente, la somma di un qualsiasi numero m con 1 (cioè $S0$) è il successore di m .

Dimostrazione.

- $m+S0$
- per la seconda definizione, $m+S0=S(m+0)$
- per la prima definizione, $m+S0=S(m+0)=Sm$

QED

Peano esprime i suoi assiomi in logica matematica, sia pur in una versione embrionale che precede i lavori di Frege. Il primissimo obiettivo di Frege, infatti, è dimostrare che gli assiomi di Peano sono derivabili dalla sua logica: poiché l'aritmetica dei numeri naturali è già ridotta agli assiomi, riducendo questi alla logica si riduce automaticamente anche l'aritmetica associata. Tarski e altri matematici hanno poi tentato un'operazione simile per la geometria euclidea, proponendo vari sistemi di assiomi logici capaci di sostituire i cinque postulati classici.

La speranza di Frege è che, andando avanti così, si arrivi a dimostrare che tutti i settori della matematica (e magari non solo della matematica) sono riducibili ad assiomi logici. Ciò che il logicismo punta a produrre, quindi, è semplicemente un sistema di assiomi non-matematici dai quali poter implicare tutte le proposizioni matematiche.

La tesi aristotelica 2: il logicismo tra Frege e Russell

Anzitutto, bisogna ridurre gli assiomi di Peano. Il solito David Hume, che aveva un'opinione su tutto, scrive: «quando due numeri sono così combinati, poiché uno ha sempre un'unità che corrisponde ad ogni unità dell'altro, li diciamo uguali». Frege, ispirandosi però a Cantor, esprime la stessa idea tramite questo assioma, oggi noto come *principio di Hume*:

PH il numero delle F è uguale al numero delle G sse tra le loro estensioni c'è *corrispondenza uno-a-uno*.

Esempio. F indica i primi ventisei successori di 0, G indica le lettere dell'alfabeto: poiché esiste una funzione che collega ad ogni F una G e ad ogni G una F (1-A, 2-B, 3-C...), il loro numero deve essere uguale.

Orbene, posto tale principio e le regole della nuova logica proposizionale da lui inventata, Frege riesce a derivare tutti gli assiomi di Peano: PH, se letto attraverso la logica fregeana, contiene già in sé tutta l'aritmetica dei numeri naturali. Un risultato strabiliante che merita mesi di festa in tutto il regno! E invece Frege non è soddisfatto: sorprendentemente, rifiuta PH, non lo accetta come assioma valido. Perché?

Ridurre la matematica alla logica richiede di dimostrare che le proposizioni matematiche sono derivabili da assiomi logici. Dunque il logicista deve partire da una serie di affermazioni *universalmente riconosciute* come parte della logica, e non da affermazioni biochimiche, sociologiche, filosofiche o quant'altro. Che una certa affermazione sia parte della logica significa, come minimo, che è analitica. Quindi PH è un principio logico solo se è analitico, solo se il concetto di *numeri uguali* ha già dentro di sé il concetto di *estensioni corrispondenti uno-a-uno* esattamente come il concetto di *scapolo* ha in sé il concetto di *uomo-non-sposato*. Oggi molti filosofi credono che il principio non soddisfi questo requisito.

Pongo F come «il numero delle Cime di Lavaredo» e G come «il numero delle persone in Dio». Il principio non dice *quali* siano questi numeri, non dice come calcolarli, non dice che debbano essere uguali al numero degli enti nelle estensioni. Io posso affermare, quindi, che entrambi questi numeri siano *il tre*. Dato ciò per vero, c'è relazione uno-a-uno tra le estensioni? Chi lo sa: se Dio esiste sì, se non esiste no (in quanto G avrà estensione vuota). Quindi la verità di PH è sintetica, dipende da quali enti esistono nel mondo, e non dal significato dei suoi termini; perciò non può essere un assioma della logica.

Il motivo per il quale Frege davvero si convince di rinunciare a PH è però un altro: il suo porre il concetto di *numero* come primitivo. Per capire la problematicità di questo fatto devo tornare agli assiomi di Peano.

Peano pone tre concetti primitivi: *numero naturale*, *successore*, e 0 . Proprio perché primitivi, non sono definiti, epperò restano *interpretabili*. Al che, io potrei capire così:

- *numero naturale* significa uno dei numeri della serie $2^{-\sqrt{x^2}}$, cioè banalmente: $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$
...;
- 0 significa 1;
- *successore* significa metà, un numero n è il successore di m sse $n \times 2 = m$.

E ora reinterpreto di conseguenza gli assiomi:

1. « 0 è un numero naturale» lo interpreto come «1 appartiene alla serie $2^{-\sqrt{x^2}}$ », ed è vero;
2. «il successore di un numero naturale è un numero naturale» lo interpreto come «la metà di un numero della serie $2^{-\sqrt{x^2}}$ è un altro numero della serie $2^{-\sqrt{x^2}}$ », ed è vero;
3. sia la funzione *successore* che la mia funzione *metà* sono iniettive, la metà di x non può essere anche la metà di un altro numero y ;

4. «per ogni numero naturale, il suo successore è diverso da 0» lo interpreto come «la metà di ogni numero della serie $2^{-\sqrt{x^2}}$ è diversa da 1», ed è vero;
5. «l'insieme di 0 e i suoi successori contiene tutti i numeri naturali» lo interpreto come «l'insieme di 1 e i suoi dimezzamenti contiene tutti i numeri della serie $2^{-\sqrt{x^2}}$ », quindi anche il principio d'induzione rimane valido.

È la stessa operazione che Putnam fa con gatti e ciliegie: posti gli assiomi di Peano, io potrei interpretarli come riguardanti l'aritmetica dei numeri naturali, mia nonna come riguardanti serie di dimezzamenti, ed entrambi i modelli rispetterebbero tutti i vincoli interni alla teoria in quanto i suoi assiomi continuano a risultare veri.

Questa ambiguità è indesiderabile: gli assiomi di Peano *incarnano* tutte le progressioni (di successori, di dimezzamenti, di punti, di istanti temporali...), quando invece dovrebbero incarnare solo le progressioni caratteristiche dell'aritmetica. La soluzione? Rifiutare i tre concetti di Peano come primitivi, bisogna invece definirli tramite altri concetti. Ma ecco il problema: anche il principio di Hume pone il concetto di numero come primitivo, quindi derivarvi gli assiomi di Peano mantiene l'ambiguità, non pone alcun vincolo che mi impedisca di interpretare i numeri come l'insieme degli imperatori romani, se così mi gira.

Dunque, come prossima mossa, Frege deve definire il concetto di numero, deve ridurre lo stesso PH ad un nuovo assioma. Pone così la sua *Basic Law V*:

BLV per ogni F e ogni G , l'estensione di F è uguale all'estensione di G sse, per ogni elemento a , « a è F » è equivalente ad « a è G »

L'assioma questa volta è analitico, perché l'estensione di F è definita da tutti gli elementi per i quali « a è F » è vera. In questo caso F e G indicano funzioni: per esempio, F potrebbe essere la funzione «padre di x », e la sua estensione sarà l'insieme di tutti gli individui (i padri) che la funzione collega ad elementi x dell'insieme di tutti i figli. L'assioma, in altre parole, afferma che se due concetti hanno un'uguale estensione, tutto ciò che è uno è anche l'altro, tutto ciò che è uno scapolo è anche un uomo-non-sposato.

Frege dimostra che da BLV si deriva PH e questo crea una definizione di *numero* davvero affascinante: il *numero delle F* (n) è l'insieme di tutti gli elementi x per i quali «ci sono n x » è vera e una delle x è F ; cioè, il *numero delle F* è l'estensione composta da tutto ciò che può stare in relazione uno-a-uno con F .

La differenza coi platonici è notevole. Per il platonista, l'estensione di *il numero delle cose infinite* (l'universo e la stupidità) include un unico ente, il numero due. Frege invece dice così: il numero delle cose infinite è identico al numero delle cose che non tornano indietro (frece e occasioni), al numero delle cose certe (morte e tasse), e in generale al numero di tutto ciò che forma una coppia, quindi tutte le coppie formano l'estensione de *il numero delle cose infinite*. In termini aristotelici: ogni coppia è una *istanziatura* del due, un esempio del due, quindi quando parlo del due parlo di tutte le coppie. Generalizzando: il numero è un insieme di insiemi, non è un ente ma si riferisce ad un'enorme combinazione di enti. *Mind-blowing*.

Questo BLV è un assioma così bello, che è davvero un peccato sia anche del tutto inaccettabile.

Anzitutto, è analitico, ma non è ancora logico. Infatti, i principi logici devono anche essere *ontologicamente neutrali*: devono limitarsi a porre una serie di simboli (x, y, \dots) e una serie di regole secondo le quali manipolarli, senza utilizzare il concetto d'esistenza; quando dico « x » intendo x e non *esiste* x , perché se questi simboli siano poi sostituiti da concetti che indicano qualcosa di esistente, o di impossibile, o anche da disegni senza alcun contenuto, è del tutto irrilevante, la logica è scienza formale perché conta solo la forma. BLV, però, non è neutrale: si appella esplicitamente all'estensione, cioè un insieme di enti esistenti, non si limita a manipolare simboli; e perciò, non può essere un assioma logico.

Cosa ancor più grave, però, è che BLV comporti delle contraddizioni: Frege aveva appena completato l'opera, mancavano pochi giorni alle stampe, ed ecco che arriva una lettera di Russell ad esporgli un *paradosso del barbiere* che fa crollare tutto il suo edificio.

I paradossi del barbiere giocano sull'autoreferenza: in città c'è un unico barbiere che tagli i capelli esclusivamente a tutti quelli che non se li tagliano da soli, dunque chi taglia i capelli al barbiere? Se se li tagliasse lui sarebbe falso che li taglia *solo* a chi non se li taglia da solo, ma se non se li tagliasse sarebbe falso che li taglia a *tutti* quelli che non se li tagliano da soli... quindi? La risposta intuitiva è che la premessa è impossibile, immaginare che il barbiere lavori così genera una contraddizione, quindi è un po' come chiedersi di che colore sarebbe la Torre Eiffel se fosse *tutta* rossa e *tutta* blu, una domanda insensata. Questa però non è la risposta di BLV, che invece finisce per creare un'incoerenza:

- P1** *principio di non-contraddizione*: o F è una F o non lo è;
- P2** pongo F e G , due concetti la cui estensione comprende tutti e soli gli insiemi che non sono membri di sé stessi;
- P3** ipotizzo che F non sia membro di sé stesso, cioè che la proposizione « F non è una F » sia vera;
- P4** quindi per definizione F è nell'estensione di G , quindi anche « F è una G » è vera;
- P5** ma per BLV, poiché F e G hanno uguale estensione, tutto ciò che è uno è anche l'altro;
- C1** quindi, se è vero che « F è una G », è vero anche che « F è una F », quindi l'ipotesi deve essere falsa;
- P6** ipotizzo allora che F sia membro di sé stesso, cioè che « F è una F » sia vera;
- P7** quindi per definizione è vero che « F non è una G »;
- P8** ma poiché l'estensione dei due concetti è uguale, per la *Basic Law V* ciò che non è nell'estensione di uno non può neanche essere nell'estensione dell'altro;
- C2** quindi è vero che « F non è una F », quindi anche questa ipotesi deve essere falsa;
- C3** è dunque falso sia che F sia una F , sia che non lo sia, ma questo contrasta con il sacrosanto principio di non-contraddizione. Paradosso!

BLV è in linea con l'insiemistica di Cantor nel non porre limiti a ciò che può costituire un insieme: le funzioni F e G possono indicare qualsiasi proprietà, e per ognuna la teoria prevede un'estensione (foss'anche vuota). Di conseguenza, come Frege scrive esplicitamente, in una funzione $F=f(x)$ la x può indicare anche la f stessa, e questa è la causa del paradosso: mentre $f(x)$ (come «padre di x ») collega un insieme x (figli) con l'insieme-estensione di f (padri), $f(f)$ (come «è un insieme f che non appartiene a f ») collega in modo autoreferente l'estensione con sé stessa. Che questo porti al

paradosso era stato previsto, in insiemistica, sia da Cantor che dal matematico Zermelo alcuni anni prima di Russell, ma Frege non ne era a conoscenza (era stato già comunicato, invece, al suo amico continentale Husserl).

Sono dunque due le teorie affette dal paradosso, insiemistica e logicismo; e sono due anche le strategie di risoluzione: o si modificano gli assiomi di partenza, o si modifica la logica che regola il ragionamento. Zermelo, lavorando sull'insiemistica, mantiene la logica standard ma modifica gli assiomi di Cantor: crea così la *teoria di Zermelo-Fraenkel* oggi unanimemente accettata. Russell invece, lavorando sul logicismo assieme a Whitehead, sceglie la via più radicale e crea direttamente una nuova logica: nascono così i monumentali *Principia Mathematica*, più di mille pagine nelle quali si pongono le basi di una *teoria dei tipi*.

Affermo: «Aaron T. Beck ha tutte le qualità del grande psicologo». Dato che non sembrano esservi circoli viziosi in questa frase, «avere tutte le qualità del grande psicologo» non può essere a sua volta una qualità del grande psicologo: «avere tutte le qualità» rappresenta un certo *tipo* di concetto, le qualità specifiche ne rappresentano un altro, e la predicazione «ha tutte le x » può accettare come x solo concetti del tipo delle qualità. Così, i *Principia* tentano di risolvere il paradosso assegnando ai vari elementi una tipologia che determini e limiti le operazioni eseguibili su di essi. È un po' come impostare il tipo di una casella Excel, se imposto «data» accetterà solo valori che rappresentano date e non numeri con la virgola; similmente, $f(x)$ sarà *etichettata* in modo tale che x possa fare da segnaposto solo per elementi di tipo diverso da quello di f .

In questo modo Russell risolve il paradosso, ma ciò non basta a rendere il suo lavoro un successo. Ogni pagina dei *Principia* è macchiata da due peccati, uno veniale e uno mortale (mortale almeno dal punto di vista del logicismo). Il peccato veniale sta nel simbolismo complesso e di difficile lettura che di certo non aiuta a seguire le lunghissime dimostrazioni dei teoremi (è noto che Russell impiega la bellezza di quattrocento pagine per arrivare a dimostrare che $1+1=2$). Il peccato mortale, invece, sta nel fatto che gli assiomi alla base dei *Principia* non sono sempre considerabili parti della logica: uno degli assiomi utilizzati per implicare il principio d'induzione, l'*assioma di riducibilità*, come ammette lo stesso Russell, non è analitico e andrebbe rifiutato. I numerosi sforzi fatti per dimostrare che il principio d'induzione può essere derivato da altri assiomi dei *Principia* senza usare la riducibilità, purtroppo, sono stati inutili. Stando così le cose, potrei concludere che il più grande successo intellettuale di Russell è stato il *non* essere riuscito a dimostrare che $1+1=2$: lui stesso in vecchiaia dirà che, trovandosi a dover scegliere quali libri salvare da un incendio, sarebbe rimasto in dubbio tra il portare con sé l'ultima copia dei *Principia* e il buttarla tra le fiamme.

Dunque Russell non riesce a salvare il progetto di Frege. Altre soluzioni? Secondo una minoranza di filosofi contemporanei, i *neo-logicisti* (tra i quali c'è anche il pluralista Crispin Wright), PH è effettivamente analitico, e non doveva essere sostituito con BLV in primo luogo. Anche se il PH può aggirare paradosso di Russell, comunque il fatto che non ponga alcun limite ai concetti che possono apparire in «il numero delle F » continua a creare problemi: risulta infatti legittimo parlare di «il numero di tutte le cose identiche a sé stesse», e qual sarebbe questo numero? Non posso rispondere «infinito»: può essere che l'universo contenga infiniti enti, e ogni ente sia identico a sé stesso, ma l'infinito non è un numero. I neo-logicisti hanno chiamato il numero di tutte le cose identiche a sé stesse *numero universale* o *anti-zero*. Nell'insiemistica di Zermelo-Fraenkel, però, non esiste un *anti-zero*, un numero di tutti gli insiemi, e se un principio filosofico collide con l'insiemistica, è la filosofia che deve fare un passo indietro.

La tesi aristotelica 3: il finitismo e il formalismo di Hilbert

Per l'aristotelismo matematico, quando dico « $1+1=2$ » esprimo una verità, cioè mi riferisco con successo ad un fatto, un fatto che però non è un'entità astratta, ma bensì concreta, esistente nello spaziotempo, e quindi fisica. Sorge un problema. Aristotele accettava l'esistenza di un *infinito potenziale*, ma non di un *infinito attuale*: è insensato immaginare che in questo universo esistano ora infiniti enti, l'infinito esiste solo nel senso che, pur quanti enti possano esservi, è sempre possibile immaginare di aggiungerne una quantità arbitrariamente grande. L'impossibilità di un infinito attuale è un'idea comune sostenuta da molti filosofi; ma se in questo universo c'è un numero finito di enti, mentre per gli assiomi di Peano i numeri sono infiniti, com'è possibile che le mie proposizioni matematiche si riferiscano ad entità fisiche? E quando parlo dei numeri transfiniti di Cantor, a cosa posso riferirmi?

In effetti, ci sono matematici che negano l'esistenza di insiemi infiniti: sono gli arcinemici di Cantor, i *finitisti*, la cui filosofia è riassunta dall'oscuro Leopold Kronecker nell'apofrosma «Dio ha creato i numeri naturali, tutto il resto è invenzione dell'uomo». Sì, i numeri naturali sono infiniti, ma il numero in sé è sempre finito, e non esiste un insieme di tutti i numeri naturali: si riconosce così un infinito potenziale (perché ogni numero ha un successore), ma nessun infinito attuale (perché non ci sono insiemi infinitamente grandi). Non è abbastanza? Allora ci sono anche gli *ultrafinitisti*, che negano direttamente che i numeri siano infiniti: esiste davvero un numero più alto di qualsiasi altro, e certi numeri *troppo grandi* sono insensati in quanto non possono riferirsi ad alcunché.

Il finitismo è una teoria *revisionista*, richiede di modificare il modo in cui si pratica la matematica, richiede di gettare nella spazzatura i fondamenti attuali della materia per costruirne di nuovi. Uno dei più grandi matematici di tutti i tempi (e anche uno di quelli col cappello più strano), David Hilbert condivideva l'immanenza anti-platonista dei finitisti, epperò desiderava anche mantenere le conoscenze acquisite lungo quattro millenni di lacrime e sudore: all'apofrosma di Kronecker oppone il suo «nessuno ci cacerà dal paradiso che Cantor ha creato», evidenziando come lui non abbia bisogno di chiamare in causa Dio. Ma: se un infinito attuale è impossibile, come posso mantenere l'insiemistica di Cantor, la quale fa amplissimo uso di entità infinite?

Se per i platonisti la matematica è quasi teologia (un'indagine non-empirica su enti astratti ma reali), per i *formalisti* come Hilbert la matematica è in essenza un gioco. Quali sono i fondamenti degli scacchi? C'è una serie di simboli (pedoni, torri, cavalli, alfieri...) e una serie di regole che spiega come manipolarli, come *trasformare* la distribuzione dei simboli in quella matrice che è la scacchiera: la torre si muove in un modo, il cavallo in un altro, posso spostare un pezzo per volta, se si sposta un pezzo bianco poi se ne sposta uno nero... ed ecco, ponendo le regole si creano tutte le partite possibili. Per la matematica posso dire lo stesso: ci sono simboli, ci sono regole per manipolarli, e questo è tutto, come non esistono enti scacchistici astratti così non esistono enti matematici.

Come posso creare infiniti giochi da tavolo ponendo nuove regole, così posso creare infinite matematiche modificando gli assiomi: per questo i formalisti sono particolarmente aperti all'applicazione di matematiche e logiche non-standard. È dunque legittimo creare una matematica senza insiemi infiniti? No, perché le regole della matematica, a differenza di quelle dei giochi, non sono del tutto arbitrarie, e non è vero che i formalisti siano incapaci di spiegare perché, nella storia dell'umanità, si è sempre favorita una matematica tra le tante possibili. Come dice Hilbert, la matematica discende dalle caratteristiche dell'intelletto, cioè dalla razionalità, e se si favorisce in modo così netto un certo sistema non è per convenzione, ma per coerenza col pensiero umano. Il

compito del formalista è analizzare i sistemi di assiomi dotati di maggior valore, e non inventare nuove folli matematiche prive di qualsiasi utilità.

Dunque, come i logicisti da lui criticati, anche Hilbert crede che la matematica sia riducibile ad una serie di assiomi. Non vuole però ridurla *a qualcos'altro*, vuole solo ridurre i settori della *matematica superiore* (come l'analisi) all'aritmetica di Peano (cioè quella dei numeri naturali). Secondo Hilbert, settori come l'analisi matematica usano un linguaggio sintetico per praticità, pongono come propri assiomi le conclusioni di ragionamenti precedenti invece di ridimostrare tutto da zero; a volerlo, però, ogni proposizione dell'analisi potrebbe essere riscritta in termini aritmetici. Ridurre la matematica superiore all'aritmetica di Peano vuol dire anche ridurre i settori che hanno a che fare con quantità infinite ad un settore che prevede solo quantità finite, cioè si riducono gli *enti ideali* (come gli insiemi di tutti i numeri, che esistono solo nella mente) agli *enti reali* (i numeri naturali): così, si accetta l'ideologia del finitismo, ma senza richiedere alcuna revisione della pratica matematica.

Hilbert, come Frege, è motivato in questa sua ricerca aristotelista quasi-finitista dalla volontà di dare fondamenta certe alla matematica, una volontà che si manifesta di continuo nel suo lavoro (aveva già creato un'assiomatizzazione per la geometria più rigorosa di quella di Euclide). Frege ha fallito nel suo intento a causa di due fattori: il primo è la difficoltà nell'assicurarsi che il sistema che stava creando fosse coerente (e in effetti non lo era); il secondo è il doversi limitare ai soli assiomi della logica, evitando quegli assiomi matematici che invece alla fine ha inserito comunque. Ma Hilbert evita queste insidie.

Anzitutto, che l'aritmetica sia coerente è un'ovvietà, e che sia riducibile ad assiomi è già stato dimostrato da Peano: l'unica cosa che resta da fare è dimostrare la coerenza dell'assiomatizzazione scelta, perché se è vero che la matematica è riducibile all'aritmetica, dimostrando la coerenza dell'aritmetica si dimostra automaticamente la coerenza della matematica tutta. Questa assiomatizzazione deve rispettare anche l'indipendenza della matematica, cioè la sua coerenza con tutte le affermazioni non-matematiche di tutti i campi qualsiasi sia il loro sviluppo: per esempio, non posso porre ad assioma affermazioni fisiche, la fisica evolve specificando sempre più le sue affermazioni e la matematica dovrebbe cambiare continuamente in funzione alle nuove scoperte. Ne consegue che neppure Hilbert può pescare assiomi a piacere da ogni materia, epperò può comunque fare una cosa che Frege non poteva, cioè può usare direttamente anche alcuni assiomi matematici.

In sintesi, il *programma di Hilbert* consiste nel creare un'assiomatizzazione per l'aritmetica di Peano che sia *consistente* e *completa*: *consistente* perché non implica alcuna contraddizione, e *completa* perché può dimostrare (cioè dedurre e *ridurre*) ogni proposizione vera della matematica.

Corollario, per questa assiomatizzazione ogni proposizione matematica è *decidibile*: è sempre possibile conoscere il valore di verità di una proposizione in quanto, o questa è dimostrabile (cioè deducibile), o lo è la sua negazione. Per questo sulla tomba di Hilbert è scritto «wir müssen wissen, wir werden wissen», «dobbiamo sapere, sapremo».

Tutto bellissimo, ma purtroppo un sistema del genere è impossibile...

I teoremi di Gödel rendono l'aristotelismo improbabile 1: idee di base

I teoremi d'incompletezza di Gödel rappresentano il risultato più eminente prodotto dalla logica matematica. Famosissimi, fonti d'ispirazione per studiosi di ogni disciplina, sono l'argomento più tecnico, ma anche il più facilmente fraintendibile, di tutta la filosofia analitica: dovrò avanzare con prudenza per evitare gli errori nei quali molti cadono.

Anzitutto, perché questi teoremi sono così importanti per la filosofia matematica? Perché dimostrano che un sistema assiomatico per l'aritmetica con le proprietà desiderate da Hilbert è irrealizzabile, e che neppure il logicismo può derivare gli enti ideali infiniti dai finiti. Questo forse non distrugge l'aristotelismo matematico, ma lo danneggia a tal punto da renderlo una visione meno credibile delle alternative.

Gödel basa la sua dimostrazione su una funzione, detta *numero di Gödel*, che assegna ad ogni insieme di simboli un numero. Per rendere l'idea di questa operazione, pur in modo simbolico e imperfettissimo, espongo un esempio informatico.

Digito in linguaggio di programmazione C++ la linea di codice `cout<<"hello world\n";`. Questi sono simboli da me comprensibili e gestibili, ma il computer non li capisce, lui pensa solo per numeri. Uso quindi un software detto *compilatore* che traduce il mio codice nel *linguaggio macchina*, cioè trasforma ogni carattere nel numero binario assegnatogli dalle convenzionali tavole ASCII: «h» diventa «1101000», «w» diventa «1110111», eccetera. Ne risulta una lunghissima serie di 0 e 1 che io decido di interpretare nell'insieme come un unico numero in base due, e a questo numero corrisponde il mio codice in C++ (per la cronaca, è un numero grandissimo anche in base dieci, superiore a $1,45 \times 10^{50}$).

Ecco, il numero di Gödel svolge un'operazione simile a quella del compilatore. Ma qual è l'utilità di una simile funzione?

Prendo una classica frase autoreferente: «questa frase non è dimostrabile». Mi chiedo: tale frase è dimostrabile oppure no? Rispondere a questa domanda vuol dire o dimostrare che è dimostrabile, o dimostrare che non è dimostrabile. Ovviamente non posso dimostrare che è dimostrabile, perché dimostrassi ciò la renderei falsa, e quindi non la dimostrerei ma la falsificherei. Epperò non posso neppure dimostrare che non è dimostrabile, perché dimostrassi ciò la dimostrerei. Quindi se tale frase sia dimostrabile oppure no è *indecidibile*, non posso produrre una risposta.

Questa frase costituisce un paradosso del barbiere, è la solita funzione che si riferisce a sé stessa, e perciò non dà prolemi a sistemi avanzati come la teoria dei tipi. Ma affinché sorga un paradosso non è necessaria un'autoreferenza di questo tipo.

Ora scrivo la frase Q « P risulta in una proposizione indimostrabile quando preceduta dalla sua citazione». Per esempio, Q mi dice che la frase «“Mercurio è un pianeta” Mercurio è un pianeta» è indimostrabile, come tutte quelle che condividono la struttura “ P ” P .

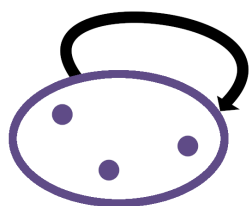
Mi chiedo: posso rendere Q una frase *direttamente autoreferente* ponendo $P = “Q”$? Posso cioè far sì che Q contenga la sua stessa citazione, e crei così il paradosso? Rappresento una frase Q che contiene varie parole a più la sua citazione così: $Q = \langle a, Q \rangle$. Data questa equivalenza, $\langle a, Q \rangle$ può essere riscritta come $\langle a, \langle a, Q \rangle \rangle$, quindi come $\langle a, \langle a, \langle a, Q \rangle \rangle \rangle$, e via così; cioè, se una frase contiene la sua citazione, deve contenere anche la citazione della citazione all'infinito, il che è assurdo. Infatti, se Q contiene le parole a , ripetendo a aggiungo la sua citazione, ma nel momento in cui lo faccio Q diventa aa , quindi è aa che devo ripetere, ma così Q diventa $aaaa$, eccetera. Dunque una frase non può contenere la sua citazione.

Epperò, ecco il *paradosso di Quine*:

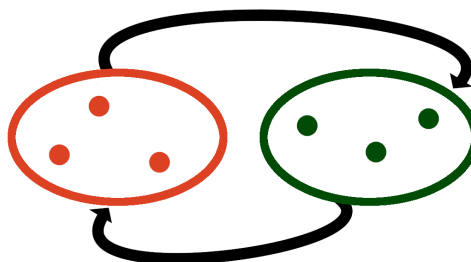
- «risulta in una proposizione indimostrabile quando preceduta dalla sua citazione» risulta in una proposizione indimostrabile quando preceduta dalla sua citazione.

Questa frase Q fa una predicazione P su un identico soggetto tra virgolette P , cioè afferma qualcosa su una sua parte ma non su sé stessa. Ciò che Q predica è l'indimostrabilità del soggetto P se preceduto da " P ", ma ovviamente " P " P risulta identico a Q . Si crea così un'*autoreferenza indiretta*: Q riesce a parlare di sé ma solo *tramite un altro fatto*, non tramite espressioni come «questa frase». È simile a «ogni generalizzazione è indimostrabile», che parla di sé solo tramite la sua classe (perché è appunto una generalizzazione).

Autoreferenza diretta



Autoreferenza indiretta



Ora, i numeri di Gödel evitano i paradossi classici rendendo l'autoreferenza diretta impossibile. Immagino di assegnare alla predicazione «è falso» il numero 693. Il numero di una frase della forma « x è falso» sarà il numero della frase-soggetto seguito da 693. Se il soggetto è la citazione di una frase, nessun problema: posta la frase 777, ottengo 777693. Ma se il soggetto è un pronome, dato che i pronomi non esistono in logica, dovrò esplicitare il numero della frase a cui si riferisce: il numero di «*quello* è falso» varia a seconda di cosa voglio indicare con «quello», se indico la frase 199 diventa 199693, se indico 784 diventa 784693, eccetera. In altre parole, nel formalismo di Gödel mi posso riferire ad altre frasi solo tramite il loro numero. Ma cosa succede se scrivo «*questa frase* è falsa»? Dovrei inserire come soggetto il numero della frase stessa, 693693, ma appena lo faccio il suo numero cambia. Lo stesso succede se provo a inserire il numero in lingua naturale: posso scrivere «693 è falsa», ma la frase «693» corrisponderà ad un diverso numero di Gödel 996, di modo che il numero sarà 996693. Quindi, come una frase non può contenere la sua citazione, così non può contenere il suo stesso numero.

Ciò nonostante, l'autoreferenza indiretta rimane possibile, e con essa i paradossi connessi. Ma questo fa parte della strategia.

Preso un'affermazione qualsiasi, coi numeri di Gödel la *aritmetizzo*, la trasformo in una proposizione aritmetica sostituendo i suoi segni con numeri. Noto così che a certe proprietà dell'affermazione corrispondono certe proprietà del suo numero: questo mi permette di risalire dal numero all'affermazione, ma mi dice anche che certe proprietà dell'affermazione vengono conservate nella sua forma aritmetizzata. Posso così dimostrare che un paradosso valido per certe formule logiche continua a valere anche per l'aritmetica, cioè che a certe strutture logiche corrispondono certe strutture aritmetiche, che alcune frasi esprimibili in logica possono essere espresse anche in termini puramente aritmetici (e viceversa).

La strategia dei teoremi di Gödel è dunque questa: si produce una frase analoga a «questa frase è indimostrabile» ma che sia indirettamente autoreferente, si dimostra che è indecidibile, e si dimostra conclude che questa indecidibilità è anche una proprietà di certe strutture aritmetiche.

Proprio per questa profonda relazione di matematica e logica, si credeva inizialmente che il particolare sistema formale utilizzato da Gödel limitasse la validità dei suoi teoremi. Per esempio, se io creo una frase indecidibile sfruttando un linguaggio capace di rappresentare non solo l'aritmetica, ma anche tutta la matematica superiore, e poi mostro che l'indecidibilità di questa frase caratterizza anche analoghe strutture matematiche, di che strutture sto parlando, esattamente? Forse, sto dicendo qualcosa di valido solo per l'analisi matematica, e non l'aritmetica. Dunque, prima di passare alle dimostrazioni, devo specificare il raggio di validità dei teoremi di Gödel.

Se prendo l'aritmetica di Peano e tolgo l'assioma d'induzione matematica ottengo l'*aritmetica di Robinson*, chiamata anche Q . Se aggiungo a questa Q un assioma d'induzione modificato in modo tale da risultare accettabile anche dal finitismo ottengo l'*aritmetica ricorsiva primitiva* (dove: se da a e b induco c , posso ripetere il ragionamento per indurre anche d , ma non posso farlo infinitamente). Queste aritmetiche non sono alternative all'aritmetica standard: è come fossero sue parti, io stesso potrei sviluppare una inutile *aritmetica di Alessio* nella quale parlo solo della somma di numeri dispari ignorando tutto il resto.

Orbene: il primo teorema di Gödel è valido per tutti i sistemi in grado di rappresentare nel proprio linguaggio l'aritmetica di Robinson, il secondo teorema invece è valido per i sistemi in grado di rappresentare nel proprio linguaggio l'aritmetica ricorsiva primitiva. Poiché entrambe queste aritmetiche sono inglobate da quella di Peano, ciò che Gödel dimostra è valido anche per quest'ultima, mentre non lo è per aritmetiche ancora più specifiche (come l'aritmetica di Alessio).

Infine, i teoremi sono validi, non per tutte le teorie consistenti, ma per tutte le teorie ω -consistenti, cioè per tutte le teorie che, oltre ad essere consistenti, se dimostrano che tutte le x hanno una certa proprietà P , non dimostrano anche che esiste una qualche generica x che non ha la tale proprietà.

Una teoria consistente non è necessariamente anche ω -consistente. Immagino una teoria capace di dimostrare sia che «0 è un numero naturale», «S0 è un numero naturale», «SS0 è un numero naturale» e via così per tutti i successori di 0, sia che «esiste un qualche successore di 0 n che non è un numero naturale», pur senza dimostrare la verità di uno specifico « n non è un numero naturale». Una teoria del genere sarebbe consistente perché non dimostra che un certo specifico numero *è e non è* un numero naturale; epperò resta una teoria indesiderabile in quanto, non essendo ω -consistente, descrive dei casi intuitivamente impossibili.

Volendo, è possibile esporre i teoremi d'incompletezza in modo tale che risultino validi anche per teorie non completamente ω -consistenti, anche se rimane necessario qualcosa in più della normale consistenza. Questa comunque non è una limitazione problematica, in quanto è desiderabile che una assiomatizzazione dell'aritmetica, oltre che consistente, sia ω -consistente.

I teoremi di Gödel rendono l'aristotelismo improbabile 2: dimostrazioni

Pongo tre simboli dell'aritmetica di Peano e assegno loro dei numeri di Gödel:

– 0: 13

- +: 17
- S: 23

Simboli in successione formano stringhe chiamate *formule*. Ogni formalismo ha una *grammatica*, una serie di regole di combinazione dei simboli: le formule che rispettano queste regole sono dette *benformate* (come «0+0» o «SS0»), le formule che invece non rispettano le regole sono dette *malformate* (come «+0S»). In ciò che segue, mi riferisco sempre a formule benformate.

Ci sono infiniti modi legittimi di assegnare un numero alle formule. Usando il metodo di Gödel, posta una formula lunga n simboli, il suo numero è il prodotto dei primi n numeri primi elevati ordinatamente al numero di Gödel dei simboli della formula. Per esempio:

$$\llbracket 0 + 0 \rrbracket = 2^{13} \times 3^{17} \times 5^{13} = 1.291.401.630.000.000.000.000$$

Perché usare una cifratura di questo tipo? Per il *teorema fondamentale dell'aritmetica*, ogni numero è rappresentabile come prodotto di numeri primi. Questo posso dimostrarlo velocemente:

- P1** suppongo che i numeri tra 1 ed n siano rappresentabili come prodotti di primi;
- P2** $n+1$ o è primo, o non lo è;
- P3** se $n+1$ è primo, $(n + 1) = (n + 1) \times 1$, quindi è rappresentabile come prodotto di primi;
- P4** se non è primo, allora per definizione $n+1$ è divisibile anche per numeri diversi da 1 e sé stesso, per esempio $\frac{n+1}{a} = b$, quindi $(n + 1) = ab$;
- P5** al che, o a e b sono numeri primi, o non lo sono;
- P6** se sono primi, allora $n+1$ è rappresentabile come il loro prodotto;
- P7** se non lo sono, allora, essendo a e b compresi tra 1 ed n , sono rappresentabili come prodotti di primi, per esempio $a = cd$ e $b = ef$;
- C** quindi $(n + 1) = cdef$, quindi è sempre rappresentabile come prodotto di primi (e poiché n può indicare qualsiasi numero, tanto 2 quanto 1231401630000000000, questa conclusione vale per ogni numero).

Ne consegue che, per tradurre un numero in formula, basta *fattorizzarlo* e disporre affiancati i simboli i cui numeri compaiono all'esponente; per esempio:

$$964.026.151.188.480.000.000.000.000 = 2^{23} \times 3^{23} \times 5^{13} = SS0$$

Non c'è rischio di ambiguità, è impossibile che ad un singolo numero corrispondano più formule; infatti:

- P1** suppongo che n sia il più piccolo numero ad avere due fattorizzazioni diverse, cioè $n = a_1 a_2 a_3 \dots$ e $n = b_1 b_2 b_3 \dots$;
- P2** suppongo che a_1 sia più piccolo di b_1 (o anche il contrario, non cambia);
- P3** esiste quindi un numero $m = (b_1 - a_1) b_2 b_3 \dots$ che deve essere più piccolo di n ;
- P4** risolvo e risulta $m = b_1 b_2 b_3 \dots - a_1 b_2 b_3 \dots$, ma essendo questa fattorizzazione equivalente a $m = n - a_1 b_2 b_3 \dots$, si dà anche $m = a_1 a_2 a_3 \dots - a_1 b_2 b_3 \dots$;

C quindi m deve avere almeno due fattorizzazioni, quindi la premessa è falsa (perché m è più piccolo di n ma avrebbe comunque due fattorizzazioni), e dato che n può indicare qualsiasi numero, non ci sono numeri con più fattorizzazioni.

Normalmente, in crittografia, si moltiplicano tra loro numeri primi molto grandi: così, per passare dal messaggio al numero basta risolvere una moltiplicazione, mentre per risalire dal numero al messaggio bisogna andare per tentativi finché si azzeccano i primi utilizzati. In questo caso, però, sapendo che i numeri di Gödel sono prodotti moltiplicando i primi numeri primi, è possibile tradurre velocemente in entrambe le direzioni.

Ogni formula ha il suo numero di Gödel, ma anche le *liste* di formule lo hanno. Per esempio, ponendo arbitrariamente l'1 come divisore, posso scrivere:

$$\langle 0 + 0; SS0 \rangle = 2^{13} \times 3^{17} \times 5^{13} \times 7^1 \times 11^{23} \times 13^{23} \times 17^{13}$$

Si tratta di un'operazione molto utile. Infatti, che cos'è la dimostrazione di un teorema matematico? È una serie di premesse, che sono formule, dalle quali deduco la proposizione del teorema, che è una formula (per esempio, una dimostrazione della proposizione b è la lista $\langle a, a \text{ implica } b, b \rangle$). Di conseguenza, per ogni dimostrazione di ogni teorema esiste un numero di Gödel.

Posso descrivere ora le *relazioni* esistenti tra le formule e tra i loro numeri di Gödel.

Le relazioni sono una generalizzazione delle funzioni: ogni funzione è una relazione, ma solo alcune relazioni sono funzioni. La differenza è questa: mentre la funzione collega un certo valore ad un unico altro, la relazione può collegare uno stesso valore anche a più valori. Per esempio, «è padre di x » è una funzione perché ogni x sarà collegata all'unico suo padre, mentre «è divisore di x » è una relazione perché uno stesso numero può avere più divisori. In ciò che segue, parlo di relazioni senza specificare se siano anche funzioni oppure no.

Le relazioni logiche continuano ad esistere anche dopo l'aritmetizzazione. Per esempio, c'è in logica una *relazione d'implicazione* che, date due formule a e b , mi restituisce $a \text{ implica } b$:

$$\text{implica}(a, b) = (a \text{ implica } b)$$

Aritmetizzo a , b e $a \text{ implica } b$ calcolando i numeri di Gödel:

- a : 2
- b : 4
- $a \text{ implica } b$: 3

Come detto, nel formalismo di Gödel posso riferirmi alle formule solo tramite il loro numero, quindi scriverò:

$$\text{implica}(SS0, SSSS0) = SSS0$$

Ma nulla distingue una formula come questa da, per esempio,

$$\text{somma}(SS0, SS0) = SSSS0$$

che è $2+2=4$ in termini logici. Volendo, posso rendere l'implicazione in stile aritmetico: come $+$ è il simbolo della relazione somma, così scelgo \rightarrow come simbolo della relazione d'implicazione e scrivo

$$2 \rightarrow 4 = 3$$

Da qui in avanti userò lettere greche per indicari numeri qualsiasi. Quando queste lettere compaiono all'interno di una formula sono sempre da interpretare in termini logici (cioè come 0 preceduto da un certo numero di \mathcal{S}).

Pongo la lista $\langle a, a \text{ implica } b, b \rangle$, che per brevità chiamo L , e calcolo i numeri di Gödel sia per la lista che per i suoi membri:

- a : α
- $a \text{ implica } b$: β
- b : γ
- L : δ

Poiché a e $a \text{ implica } b$ sono le premesse dalle quali la conclusione b discende, individuo un relazione R tale che, dandole in pasto le premesse, mi restituisce tale conclusione (è cioè una *regola di deduzione*):

$$R(\alpha, \beta) = \gamma$$

che in lingua naturale si rende con

« b è deducibile da a e a implica b »

Se tra gli elementi di una lista si dà la relazione R , allora si dà anche una *relazione di dimostrabilità* P tra il teorema e la sua dimostrazione, che in questo caso sono rispettivamente b e la lista tutta:

$$P(\gamma) = \delta$$

« L è una dimostrazione di b »

Si tratta di una relazione importante perché alla risolvibilità dell'operazione aritmetica corrisponde la dimostrabilità della proposizione logica.

Pongo ora una formula qualsiasi che contenga una variabile x : la chiamo $A(x)$, e il suo numero è ϵ . Per il *teorema di rappresentabilità* (che non dimostro) posso sostituire le x di $A(x)$ con l'equivalente logico di ϵ e creare così la nuova formula $A(\epsilon)$ di numero η . Anche questa operazione può essere espressa tramite una *relazione di sostituzione* Q che, dati due ϵ (uno come formula, uno come numero), restituisce $A(\epsilon)$:

$$Q(\epsilon, \epsilon) = A(\epsilon)$$

Adesso, anche le relazioni come $Q(\epsilon, \epsilon) = A(\epsilon)$ sono formule logiche, e dato che $A(x)$ indica una formula qualsiasi essa stessa può essere una relazione. Ne consegue che le relazioni: hanno un numero di Gödel, possono relazionare altre relazioni, e possono contenere delle variabili. Costruisco dunque una lista contenente una relazione di sostituzione nella quale sia la formula di partenza, sia il valore da sostituire, sono variabili x , chiamo tale lista $D(x)$ e il suo numero è θ :

- $D(x)$: $\langle A(y); Q(x, x) = y \rangle$
- $D(x)$: θ

Dunque sostituisco alla prima x la medesima $D(x)$, e alla seconda x l'equivalente logico del numero θ ; creo così una nuova lista $D(\theta)$ di numero κ :

- $D(\theta)$: $\langle A(y); Q(\theta, \theta) = y \rangle$
- $D(\theta)$: κ

Mi chiedo: cosa rappresentano y e $A(y)$? y è il risultato di $Q(\theta, \theta)$, ma $Q(\theta, \theta)$ è l'operazione che ha prodotto $D(\theta)$, quindi y deve essere il numero κ , mentre $A(y)$ è una formula che lo contiene:

$$D(\theta): \langle A\kappa, Q(\theta, \theta) = \kappa \rangle$$

Questo non vuol dire che il numero di Gödel di y sia κ : il numero κ scritto in termini logici avrà come numero di Gödel un diverso ι . Dato che non esistono più numeri con una identica fattorizzazione, non esistono neanche più formule con lo stesso numero di Gödel.

Ora posso facilmente dedurre il *lemma di diagonalizzazione*:

- P1** pongo la lista $D(\theta) \langle A(y); Q(\theta, \theta) = y \rangle$ di numero κ ;
- P2** poiché $Q(\theta, \theta) = y$ produce $D(\theta)$, $Q(\theta, \theta) = y$ e $y = \kappa$ sono equivalenti, infatti:
- P2.1** se $Q(\theta, \theta) = y$ è vera, producendo $D(\theta)$, y deve essere il suo numero, cioè κ ;
- P2.2** se invece $Q(\theta, \theta) = y$ è falsa, producendo $D(\theta)$ cioè κ , è impossibile che y sia κ (lo fosse, sarebbe vera);
- P3** quindi la formula $D(\theta)$ può essere riscritta come $\langle A(y); y = \kappa \rangle$;
- C** ma allora $D(\theta)$ è equivalente ad $A(\kappa)$, infatti:
- C.1** se $D(\theta)$ è vera, entrambi gli elementi della lista sono veri, quindi si dà $A(\kappa)$;
- C.2** se invece $D(\theta)$ è falsa:
- C.2.1** o entrambi gli elementi della lista sono falsi, e quindi anche $A(\kappa)$ deve esserlo;
- C.2.2** o non si dà $A(y)$, e quindi neanche $A(\kappa)$ benché $y = \kappa$;
- C.2.3** o non si dà $y = \kappa$, e quindi neanche $A(\kappa)$ benché $A(y)$.

Poiché $D(\theta)$ e $A(\kappa)$ indicano formule qualsiasi, il teorema è di validità generale. Cioè, posta una certa formula con incognita $A(x)$, e definito \leftrightarrow come simbolo dell'equivalenza logica, esisterà una certa formula D tale che:

$$D \leftrightarrow A(d)$$

Ed ecco il trucco dell'autoreferenza: D è equivalente ad una formula $A(d)$ che contiene il suo numero.

Decido quindi di porre come $A(x)$ la negazione della relazione di dimostrabilità P , $non-P(x)$. Dunque, posto n come qualsivoglia formula possibile, per il lemma di diagonalizzazione esisterà una qualche formula D di numero d equivalente a

$$non P(d) = n$$

«non vi è dimostrazione di D »

Dunque, ecco il *primo teorema d'incompletezza*:

- P1** ipotizzo che D sia dimostrabile;
- P1.1** quindi esiste una lista di formule L rappresentante la prova di D , cioè si dà il caso che $P(d) = L$;
- P1.2** ma D è equivalente a $non-P(d) = n$, quindi, se si dà il caso di $P(d) = L$, L prova anche $non-P(d) = n$;

- P1.3** il sistema, se consistente, non può dimostrare sia $P(d)=L$ che $non-P(d)=n$, quindi la premessa deve essere falsa;
- P2** ipotizzo allora che la negazione di D , $non-D$, sia dimostrabile;
- P2.1** se esiste una prova L di $non-D$, e il sistema è consistente, allora non può esistere alcuna prova di D ;
- P2.2** quindi il sistema può dimostrare che $non-P(d)=n$, cioè il sistema dimostra che qualsiasi formula n ha la proprietà di essere una *non-dimostrazione* di D ;
- P2.3** ma dato che $non-D$ è equivalente a $P(d)=m$ (dove m è una qualche formula), se $non-D$ è dimostrabile, allora è dimostrabile anche che esiste una qualche m tale che $P(d)=m$;
- P2.4** il sistema, se ω -consistente, non può dimostrare contemporaneamente $non-P(d)=n$ e l'esistenza di una m tale che $P(d)=m$;
- P2.5** quindi il sistema non può dimostrare $P(d)=m$, quindi non può neanche dimostrare l'equivalente $non-D$, quindi la premessa deve essere falsa;
- C** quindi, in un sistema ω -consistente, D è una *frase di Gödel*, cioè è indecidibile.

Il *secondo teorema d'incompletezza* segue intuitivamente, anche se la dimostrazione originale è invero lunga e complessa.

Chiamo F (con numero f) una qualsiasi formula autocontraddittoria: un esempio classico è $0=S0$, ma non ci sono limiti alle contraddizioni rappresentabili coi numeri di Gödel. Indico con n tutte le formule possibili e con m una qualche formula generica, quindi scrivo:

$$P(non\ P(f) = n) = m$$

« m è una dimostrazione del fatto che non vi è dimostrazione di F »

Se un sistema formale può produrre una dimostrazione del fatto che non può produrre dimostrazioni di F , potendo essere F una qualsiasi contraddizione, quel sistema formale dimostra la sua propria consistenza. Dunque:

- P1** ipotizzo che il sistema formale utilizzato possa dimostrare la sua stessa consistenza, cioè che possa produrre una m tale che $P(non-P(f)=n)=m$;
- P2** come dimostrato nel primo teorema, se un sistema è consistente, allora D non è dimostrabile in quel sistema, cioè si dà $non-P(d)=n$ dove n è qualsivoglia formula possibile;
- P3** quindi, se il sistema può dimostrare la sua stessa consistenza, può dimostrare anche ciò che tale consistenza implica, cioè $non-P(d)=n$;
- P4** ma $non-P(d)=n$ è equivalente a D , quindi se il sistema può dimostrare $non-P(d)=n$, può dimostrare anche che esiste una qualche prova k tale che $P(d)=k$, il che contraddice il primo teorema ed è inaccettabile in un sistema ω -consistente;
- C** quindi, se il sistema è ω -consistente, non può dimostrare la sua stessa consistenza.

Ciò non significa, come alcuni intendono, che ogni sistema contenga contraddizioni, o che esistano verità in linea di principio inconoscibili: in effetti, un sistema incapace di rappresentare l'aritmetica di Robinson può nondimeno essere sia consistente che capace di dimostrare ogni sua verità (inclusa la sua consistenza). Inoltre, anche se il mio sistema non è in grado di dimostrare certe verità, un secondo sistema potrebbe essere in grado di farlo, e in tal caso io posso comunque conoscere ogni verità esprimibile dal mio sistema.

Ora, le ovvie implicazioni filosofiche. Hilbert vuole creare un'assiomatizzazione che dimostri tutte le verità matematiche senza dimostrare alcuna contraddizione, sogna cioè un sistema sia consistente che completo. Ma per il primo teorema d'incompletezza, se questo sistema è ω -consistente, allora, dovendo comprendere una proposizione D indecidibile, è incompleto.

Sollevo un'obiezione: questa proposizione indecidibile D potrebbe anche essere falsa, e se lo è non costituisce affatto un problema per il programma di Hilbert, il quale punta a dimostrare solo le affermazioni vere. Peccato però che D non possa essere falsa:

- P1** D è equivalente a $\text{non-}P(d)=n$, cioè se D è vera allora ogni formula è una *non-prova* di D ;
P2 se D è falsa, allora deve esistere un numero m tale che $P(d)=m$ (se non esiste, non è falsa);
C ma per i teoremi d'incompletezza questo numero m non può esistere, quindi D non può essere falsa.

D è una proposizione vera ma indimostrabile: è vero che $\text{non-}P(d)=n$ ma non esiste una m tale che $P(\text{non-}P(d))=m$. Dunque, posto un sistema ω -consistente che rappresenta l'aritmetica di Peano, esiste qualche verità che è indimostrabile in tale sistema: perciò, non solo tale sistema non può dimostrare tutte le verità matematiche, ma non può neppure dimostrare tutte le verità aritmetiche. Per di più, per il secondo teorema d'incompletezza, contrariamente alle speranze di Hilbert, dimostrando la consistenza dell'aritmetica tramite l'aritmetica stessa non si dimostra la consistenza della matematica tutta, ma al contrario si dimostra la sua inconsistenza.

E così la tesi di Hilbert viene ribaltata: l'aritmetica non può dimostrare ogni verità della matematica, al contrario è la matematica superiore che dimostra alcune verità aritmetiche altrimenti indimostrabili. Di conseguenza, i programmi riduzionisti risultano indesiderabili: i vari settori matematici devono mantenere autonomia perché solo così potranno supportarsi l'un l'altro nel dimostrare le proprie proposizioni.

I teoremi sono dunque devastanti per il formalismo, ma hanno effetti di rilievo anche sul logicismo.

Frege, Russell, e perfino Zermelo esprimono le proprie teorie in linguaggi equivalenti alla logica del primo ordine o ad una logica di ordine superiore. Se questi linguaggi sono ω -consistenti e completi, allora non hanno i mezzi per rappresentare l'aritmetica di Peano; ma se possono rappresentarla, o non sono davvero ω -consistenti (non sia mai!), o non implicano davvero ogni verità matematica. Un esempio insiemistico che dimostra quanto detto è l'*ipotesi del continuo*. Posta la cardinalità dei numeri naturali \aleph_0 e la cardinalità dei numeri reali \mathfrak{c} , «non esiste alcun insieme la cui cardinalità è tra \aleph_0 e \mathfrak{c} »: questa proposizione insiemistica o è vera o è falsa, ma è indecidibile all'interno dell'insiemistica di Zermelo-Fraenkel, cioè la loro teoria non implica né questa ipotesi né la sua negazione.

Questo non nega la possibilità di creare una combinazione di diversi sistemi logici che, nel complesso, implichi ogni verità matematica. Ma se il logicismo si accontentasse di questo, sarebbe una teoria fin troppo modesta. La tesi del logicismo è che le verità matematiche siano tali in virtù di una qualche struttura logica. Io posso costruire sistemi logici capaci di implicare qualsiasi proposizione, anche tutte le proposizioni fisiche, ma questo non mi permetterebbe certo di ridurre la fisica alla logica: le proposizioni fisiche sono rese vere dal corrispondere ai fatti fisici, non dalla loro derivabilità da certi assiomi logici. Similmente, il solo fatto che ogni verità matematica sia

derivabile da qualche sistema logico non mi permette di difendere la tesi logicista, non mi permette di concludere che la logica è il *truthmaker* delle proposizioni matematiche.

I teoremi di Gödel rendono l'aristotelismo improbabile 3: altre implicazioni filosofiche

Douglas Hofstadter è, secondo me, l'unico genio vivente. Figlio del fisico Hofstadter (che fu ispirazione per il Leonard di Big Bang Theory), è anzitutto filosofo della mente, ma anche pianista, disegnatore di *ambigrammi* (immagini ambigue), e pure poliglotta (sua moglie era un'italianista, tutt'ora nei rapporti famigliari predilige l'italiano).

Il suo argomento preferito è l'autoreferenza indiretta. Come dottorando in fisica, ha scoperto la *farfalla di Hofstadter*, un frattale, cioè un'immagine internamente omeotetica (che contiene sé stessa su infinite scale di grandezza) utile a descrivere certe proprietà degli elettroni: già in questo c'è autoreferenza. Nel suo lavoro filosofico più famoso invece, «Gödel, Escher, Bach», Hofstadter analizza il concetto di autoreferenza indiretta tramite l'arte per giungere a conclusioni sulla natura della mente che sono a cavallo tra la biologia e l'informatica.

Hofstadter espone, ma non sposa, questo semplice ragionamento:

- P1** se la mente umana è un meccanismo (come un computer), allora, *se è consistente*, è soggetta ai limiti indicati dai teoremi di Gödel;
- P2** ma, diversamente da quanto suggerirebbero i teoremi, la mente umana può intuire la verità di certe proposizioni indecidibili, il matematico professionista intuisce se un'equazione ha una soluzione anche prima di mettersi a calcolare;
- C** quindi la mente umana non è un meccanismo.

Gödel stesso aveva già considerato questa possibile interpretazione dei suoi teoremi, ma afferma che è ben possibile che l'uomo sia impossibilitato a dimostrare la verità di certe equazioni. Putnam, invece, rifiuta la conclusione in quanto la mente non è consistente, l'essere umano si contraddice di continuo e per questo può essere in grado di dimostrare ogni proposizione. John Lucas, un filosofo che quando ho iniziato a scrivere queste righe era ancora in vita (ora è in cielo), ha sviluppato una versione più raffinata di questa argomentazione:

- P1** ipotizzo che l'uomo sia una macchina determinata;
- P2** se è così, allora esiste un sistema logico deterministico L in grado di prevedere ogni azione umana in ogni circostanza;
- P3** per i teoremi d'incompletezza, è sempre possibile costruire all'interno di un sistema logico abbastanza complesso una formula vera ma indimostrabile in quel sistema;
- P4** di conseguenza un matematico (usando un computer di potenza illimitata, se serve) può costruire una formula F tale da essere vera ma indimostrabile all'interno del sistema L che prevede tutte le sue azioni;
- P5** ma il matematico, in base al metodo utilizzato per costruire la formula, sa che F è vera;
- P6** quindi il sistema L non può prevedere tutte le azioni del matematico (il matematico può dimostrare la formula, il sistema L no);
- C** quindi la premessa deve essere falsa, l'uomo non è una macchina determinata.

Scienziati e filosofi eminenti come Penrose, Nagel, Searle e Crispin Wright hanno supportato argomentazioni simili a questa, la quale però continua ad essere soggetta all'obiezione sollevata dallo stesso Gödel.

Adesso provo a modificare il ragionamento in modo tale che, invece di difendere il libero arbitrio, giustifichi il platonismo:

- P1** se la matematica è una creazione della mente umana, allora l'uomo è in grado di risolvere ogni equazione (in quanto non può creare equazioni le cui soluzioni sono al di là di quelle che può formulare);
- P2** ipotizzo che la matematica sia una costruzione della mente umana, e quindi che l'uomo sia capace di risolvere ogni equazione;
- P3** per i teoremi di Gödel, nessun sistema sufficientemente complesso può dimostrare ogni sua proposizione, e quindi, se la mente può risolvere ogni equazione, deve avere capacità di calcolo superiori a quelle di qualsiasi macchina possibile;
- P4** questa conclusione è anti-naturalista e scientificamente inaccettabile, quindi la premessa deve essere falsa;
- C** quindi, essendo la mente un meccanismo, devono esistere equazioni irrisolvibili dalla mente umana o da qualsiasi altra macchina, quindi la matematica non può essere una costruzione della mente umana e deve avere invece esistenza fattuale.

Si tratta di un'argomentazione probabilistica: senza dubbio, un informatico può creare sistemi che superano le capacità di qualsiasi uomo incluso sé stesso, i moderni motori scacchistici sono più forti di qualsiasi giocatore umano. Certo, un computer è frutto di un lavoro collettivo, non dell'opera di un uomo solo, ma nulla vieta in linea di principio che un uomo solo possa costruire un computer del genere, avendo a disposizione abbastanza tempo.

La tesi kantiana: Brouwer e l'intuizionismo

Kant non era solo un filosofo: come il fisico-linguista-musicologo-egittologo Thomas Young, il biogeografo ed esploratore Alexander von Humboldt, il fisico e medico von Helmholtz, anche Kant è a volte presentato quale ultimo genio universale, l'ultimo uomo a conoscere tutto ciò che era rilevante conoscere nella sua epoca. Gli interessi di Kant infatti spaziavano anche nelle scienze naturali e formali: la matematica l'aveva insegnata come professore, aveva intuito la possibile esistenza dei buchi neri, e contemporaneamente a Laplace aveva teorizzato l'origine del sistema solare da una nebulosa.

La filosofia matematica di Kant è poco nota ma di notevole importanza.

Secondo Kant il filosofo procede per *analisi*, spezzetta i concetti in componenti essenziali; il matematico, invece, procede per *sintesi*, parte da concetti minimi certi e li combina per *costruire* qualcosa di nuovo: partendo dai concetti di linea e angolo posso costruire il quadrato. La matematica però ha uno statuto particolare: mentre le altre materie sintetiche (come la fisica) costruiscono conoscenza a posteriori, la matematica funziona a priori.

Ovviamente vi sono delle obiezioni: secondo Quine e Putnam, per esempio, anche in matematica vi è conoscenza sperimentale a posteriori, non solo a priori; Frege e Russell, invece, accettano che la matematica sia totalmente a priori, ma non possono accettare che sia anche sintetica (in quanto deve essere riducibile a proposizioni analitiche). Ma Kant pare avere ragione: che sia sintetica è comprensibile alla luce dei teoremi di Gödel, dato che la conoscenza degli assiomi non è sufficiente a implicare la verità di certe proposizioni; e che sia a priori, *be'*, è un'intuizione comune derivata dall'osservazione della propria pratica matematica.

È vero, è una strana combinazione di proprietà: se un'affermazione è sintetica, cioè non posso provare la sua verità in virtù del significato dei termini, come posso conoscerla a priori? Kant, osservando il lavoro dei matematici della sua epoca, scopre che molte dimostrazioni di teoremi non sono affatto rigorose, si basano invece su visualizzazioni mentali o altro tipo di rappresentazioni simboliche dei concetti matematici. Conclude così che le verità matematiche sono inferite sulla base di *intuizioni*. Un geometra non costruisce il triangolo scrivendo che è «la figura racchiusa in tre rette incidenti una con l'altra», lo costruisce mostrando la figura (almeno nell'immaginazione) e solo dopo la definisce così. Anche se un bambino possiede i concetti di 5, 7 e *somma*, analizzandoli non arriverà mai al concetto di 12, è l'intuizione sintetica che nella mano ci sono cinque dita una di fianco all'altra a permettergli di vedere come il dodici emerge da tali concezioni. Certo, Peano mostra che in effetti il 12 è derivabile anche da assiomi puramente logici, da un'analisi puramente concettuale; non pare però vero, come vuole Russell, che l'emersione di settori *controintuitivi* della matematica, come la geometria non-euclidea, non lasci più spazio all'intuizione, anzi: l'intuizione può comunque essere l'origine degli assiomi stessi.

La differenza con l'aristotelismo diventa così evidente: per i logicisti, una volta posti gli assiomi, tutti i teoremi della matematica esistono già, io devo solo scoprirli derivando proposizioni; ma nel kantismo, dato che la matematica non è solo ragionamento ma anche intuizione, la derivabilità logica di certi teoremi non è sufficiente alla loro verità, essi esistono solo quando li *costruisco* con l'intuizione. In altre parole, il kantismo matematico prevede questi tre elementi:

1. *costruttivismo*: gli enti matematici sono costruiti tramite l'attività del pensiero, ad esempio tramite sintesi di concetti semplici;
2. *intuizionismo*: questi concetti semplici sono frutto di intuizioni matematiche;
3. *antirealismo*: quindi gli enti matematici non sono reali, non sono riducibili, né sono simboli di qualcosa di reale, sono pure e semplici costruzioni.

Henri Poincaré è senza dubbio il più grande tra i matematici influenzati da Kant: benché al suo tempo non fosse più possibile avere dominio di molteplici campi di studio (come poteva invece fare Kant), Poincaré si è dimostrato comunque un genio universale capace di dare contributi fondamentali in ogni settore della matematica, inclusi quelli applicati come fisica e ingegneria. Avversario di Cantor ma anche del logicismo, Poincaré oppone a Russell l'idea che l'intuizione sia il cuore dell'attività matematica, una conclusione alla quale era giunto indagando l'attività del suo stesso cervello al lavoro.

Contro il logicismo, Poincaré offre una possibile dimostrazione della natura sintetica della matematica. Posto che «se k ha una certa proprietà allora ce l'ha anche il suo successore», per l'induzione matematica posso concludere direttamente che tutti i numeri hanno quella proprietà. Ma come faccio a verificare quanto detto, a verificare che non vi siano numeri per i quali l'induzione non vale? Dovrei controllare di numero in numero se la proprietà è conservata, cioè dovrei farlo con metodo induttivo. Così, l'induzione matematica può essere dimostrata solo come una proposizione sintetica, non analitica, va inferita tramite osservazioni ripetute e non tramite

l'analisi dei termini. Anche Wittgenstein, pur sfuggendo ogni classificazione, aveva idee simili a questa, mentre Frege, anche lui influenzato da Kant ma a favore di una diversa nozione di sinteticità, respinge l'argomentazione (poteva riconoscere come sintetica a priori la geometria, ma non l'aritmetica).

L'interpretazione più fedele e radicale della tesi kantiana si trova in L. E. J. Brouwer, l'*enfant terrible* della matematica, prima amico e poi nemico giurato di Hilbert (il quale da amico gli diede un posto di lavoro, e da nemico glielo tolse).

L'idea di una matematica totalmente costruita, opposta alla teoria platonica di un regno in cui i numeri astratti esistono davvero, può sembrare materialista, scienziata; in realtà Brouwer aveva una visione della vita mistica paragonabile a quella dei platonici, sosteneva un volontarismo orienteggiante ispirato a Schopenhauer (Brouwer 2015). Come quest'ultimo, Brouwer era convinto che i concetti si basassero su intuizioni sensoriali, e come Kant credeva che lo spazio e il tempo fossero le prime e più pure intuizioni: vedo un oggetto, il tempo passa, e magicamente appaiono due oggetti, uno presente e uno passato, uno davanti a me e uno nella memoria, e così creo il concetto del numero 2, e poi di ogni altro numero.

Il misticismo di Brouwer emerge anche nell'immagine del *matematico ideale*. Pensa ad un matematico di abilità enormi ma finite, incapace per esempio di calcolare un numero infinito, ma in grado di dimostrare tutto e solo ciò che è vero, e tale che non vi sia nulla di vero che non possa prima o poi dimostrare. Questo matematico non è Dio, ma un *soggetto creatore*, è un'astrazione dell'umanità depurata da tutti i processi psicologici non essenziali all'attività matematica, l'umanità che col suo pensiero costruisce la matematica. Questa immagine permette di capire varie cose sul modo in cui Brouwer intende la matematica.

Da antirealista, sposa per la matematica un principio di conoscibilità: essendo la matematica costruita dalle intuizioni dell'umanità, ed essendo quindi almeno in parte soggettiva, non può esservi una verità matematica indimostrabile. Per esempio, io posso affermare l'esistenza di *numeri primi sexy*, coppie di primi la cui differenza è 6: io costruisco tale classe di entità *producendo* un loro membro, quindi mostrando per esempio la coppia {5, 11}; così, nel momento stesso in cui *creo* i numeri primi sexy, ho dimostrato la loro esistenza, ma prima che li costruissi non esistevano. Da tale idea discende il misticheggiante *principio di carità cristiana* per il quale se una proposizione è riconosciuta come indimostrabile, lo è perché riconosciuta come assurda, e non perché l'umanità sia troppo stupida per dimostrarla.

Questa visione richiede una teoria della verità non-corrisponentista per la matematica: se gli enti matematici sono costruiti, la verità delle affermazioni matematiche non può stare nella corrispondenza ai fatti, non può stare in qualcosa di indipendente dalla mente, sta solo nella mia capacità di costruire ciò di cui parlo. Così, quando io dico che $2+2=4$ «è vero», dovrei intendere, più correttamente, «è costruibile», o «è dimostrabile»: e questo è proprio ciò che intendevano i positivisti logici con il loro principio di verifica.

Come visto in epistemologia, però, i principi di conoscibilità e verifica tendono a produrre paradossi nella logica classica. E infatti Brouwer era convinto che la logica classica, sviluppata da Aristotele e mai evolutasi, non fosse in grado di rendere merito di certe caratteristiche del mondo, e che non fosse neanche adatta all'attività matematica. In fondo, la filosofia orientale ha sempre favorito alcuni principi diversi da quelli aristotelici, in India e in Cina la matematica si sarebbe potuta sviluppare seguendo un approccio diverso rispetto a quello assiomatico di Euclide, e anche

in Europa filosofi romantici come Hegel hanno preferito sposare principi logici non-classici. Ciò che non convinceva Brouwer in particolare era la *legge del terzo escluso*:

Terzo escluso: $\neg(P \wedge \neg P)$

cioè l'idea che o una proposizione è vera o lo è la sua negazione, o è vero che un liquido è acqua o è vero che non è acqua.

Per capire perché tale principio sia inaccettabile nell'approccio kantiano posso confrontare il diverso modo in cui logici classici (che lo accettano) e costruttivisti intuizionisti (che lo rifiutano) interpretano le formule. Prendo come esempio il terzo escluso: $\neg(P \wedge \neg P)$ può essere letto «o è vero P o lo è la sua negazione», ed è cosa ovvia, ma dalla prospettiva costruttivista si legge invece «o è dimostrabile P , o è dimostrabile la sua negazione», che invece è platealmente falso (dato che possono essere entrambi indimostrabili). Altro esempio equivalente: $\neg\neg P \rightarrow P$ è interpretabile come «se è falso che P è falso, allora è vero», ma nel costruttivismo diventa «se è indimostrabile che P è indimostrabile, allora è dimostrabile», e anche questa è una totale falsità. Di fatto, si sostituisce la verità giustificata con la sola giustificazione.

Il rifiuto del terzo escluso può sembrare un cambiamento da poco per la matematica. E invece è devastante. A differenza dell'approccio di Hilbert, il costruttivismo non è *conservativo*: alcuni teoremi certi della matematica classica qui non sono più validi, mentre lo sono altri principi normalmente negati.

Anzitutto, diventa impossibile offrire dimostrazioni *per assurdo*. Quando ho discusso qui sopra il teorema fondamentale dell'aritmetica ho scelto di dimostrarlo così: invece di provare direttamente la sua verità, ho dimostrato che se fosse falso si genererebbe una contraddizione, cioè ho dimostrato la falsità della sua negazione, e da ciò, tramite il terzo escluso, ho dedotto la sua verità. Ma in una filosofia costruttivista questo approccio è inaccettabile: se è vero che gli enti matematici sono costruiti, il mio dimostrare che l'inesistenza di numeri primi sexy genererebbe paradossi è insufficiente al dimostrare la loro esistenza, per farli esistere devo costruirli, credere diversamente vorrebbe dire riconoscere loro un'esistenza autonoma e quindi sposare una forma di platonismo.

Per quel che riguarda l'aritmetica, devo reinterpretare gli assiomi di Peano in logica intuizionistica, formando così l'aritmetica di Heyting. L'analisi pure viene modificata in modo costruttivista, creando una teoria per la quale alcuni teoremi dell'analisi classica non risultano formulabili in modo equivalente. E per l'insiemistica, posso dire la stessa cosa. Come esempio, prendo il *teorema di Schröder-Bernstein*, secondo il quale, posti gli insiemi A e B , le funzioni *iniettive* f e g , e la funzione *biiettiva* w , se si danno $f(A,B)$ e $g(B,A)$, allora si dà anche $w(A,B)$: cioè, se c'è una funzione f che assegna a tutti gli elementi di A degli elementi di B (non necessariamente tutti), e c'è una funzione g che assegna a tutti gli elementi di B degli elementi di A (non necessariamente tutti), allora tra A e B vi è relazione uno-a-uno. Orbene: questo teorema implica il terzo escluso, quindi non è dimostrabile nell'intuizionismo.

Il rifiuto del terzo escluso, infine, si ripercuote anche in settori esterni alla matematica: gli informatici, per esempio, sfruttano l'intuizionismo per analizzare i problemi di decidibilità (dove conta più la dimostrabilità in un tempo finito che la verità), mentre Michael Dummett, da antirealista parziale qual era, ha usato la logica intuizionistica per offrire una nuova prospettiva su vari problemi filosofici classici.

Insomma: se si accetta che la matematica è nulla più che una costruzione, allora bisogna cambiare il modo in cui la si pratica. E questo è indesiderabile.

Detto tra parentesi: un'argomentazione di Benacerraf a favore del kantismo

John von Neumann, padre del computer, è stato uno dei più originali matematici di sempre, forse tra tutti il più dotato intellettualmente: coloro che lo incontravano, inclusi i più eminenti premi Nobel del tempo, lo riconoscevano come l'uomo più intelligente che avessero mai visto, lo descrivevano come una sorta di alieno o di stadio successivo dell'evoluzione umana. Era ossessionato dal pensiero, in particolare dal pensiero matematico, e ciò gli aveva donato una memoria e una velocità di calcolo al di là di ogni immaginazione (pare che già da bambino riuscisse a memorizzare una pagina dell'elenco telefonico in pochi secondi). Appassionato di storia bizantina e di marce militari tedesche (che ascoltava a tutto volume per disturbare Einstein, che lavorava nell'ufficio di fianco al suo), pessimo guidatore ma dotato di un notevole senso dello stile, è famoso anche per la sua conversione religiosa sul letto di morte, un aneddoto spesso citato dagli apologeti.

Inizialmente interessato al programma di Hilbert, von Neumann inizia a dubitare del programma ascoltando una conferenza di Gödel a Königsberg (la città natale di Hilbert, ma anche di Kant): ne rimane così profondamente colpito che finisce per scoprire per conto suo, con solo alcuni giorni di ritardo rispetto a Gödel, il secondo teorema d'incompletezza. Il rifiuto di una prospettiva aristotelica però non deve spingere automaticamente al platonismo: proprio il lavoro di von Neumann sui *numeri ordinali*, infatti, offre una (imprevista) difesa del kantismo matematico.

Mentre i numeri cardinali descrivono la quantità di elementi contenuti in un insieme, i numeri ordinali etichettano e appunto ordinano questi elementi. Nel caso di insiemi finiti, cardinali e ordinali coincidono: per esempio, l'insieme della mia famiglia ha quattro elementi, quindi il suo cardinale è 4, e 4 è anche l'ordinale assegnato al mio cane (ordinando gli elementi per data di nascita). Nel caso di insiemi infiniti, però, non è così, bisogna distinguere cardinalità e ordinalità, dimensione e ordine; e come vi sono cardinali infiniti (e.g. \aleph_0), così vi sono ordinali infiniti (il primo numerabile è ω , il primo non numerabile è Ω).

Gli ordinali sono definiti tramite insiemi: l'idea è che un numero sia minore di un altro se il primo è un sottoinsieme del secondo. Nello specifico, von Neumann li definisce così:

$$\begin{aligned} 0 &= \emptyset \\ 1 &= \{\emptyset\} \\ 2 &= \{\emptyset, \{\emptyset\}\} \\ 3 &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \end{aligned}$$

Zermelo però offre un'altra definizione:

$$\begin{aligned} 0 &= \emptyset \\ 1 &= \{\emptyset\} \\ 2 &= \{\{\emptyset\}\} \\ 3 &= \{\{\{\emptyset\}\}\} \end{aligned}$$

Qual è la differenza? In von Neumann un ordinale è un insieme che contiene tutti i suoi predecessori: il 3 è l'insieme che contiene lo 0, l'1 e il 2 (il quale contiene lo 0 e l'1, il quale contiene

lo 0). In Zermelo, invece, un ordinale è un insieme che contiene solo l'immediato predecessore: il 3 contiene solo il 2 (il quale contiene l'1, il quale contiene lo 0). Anche se entrambe le definizioni sono giustificate e pressoché equivalenti per tutte le applicazioni aritmetiche, le risposte che danno a certe domande varia. Esempio: l'uno appartiene al tre? Von Neumann risponde di sì, Zermelo di no. Esiste un ordinale infinito ω coincidente con l'insieme di tutti i numeri naturali? La definizione di von Neumann lo prevede, quella di Zermelo invece descrive solo ordinali finiti.

Secondo l'antirealista Benacerraf questo fa sorgere dei problemi per i realisti matematici. Immagino di sostenere una versione minimale del platonismo, un *platonismo degli insiemi* per il quale gli unici enti matematici esistenti sono quelli fondamentali. Ora mi pare corretto sostenere che $2 = \{\emptyset, \{\emptyset\}\}$, ma anche $2 = \{\{\emptyset\}\}$: infatti, entrambe le definizioni rendono merito delle medesime verità aritmetiche. Da ciò deduco che $\{\emptyset, \{\emptyset\}\} = \{\{\emptyset\}\}$, ma non può essere così: se è vero che gli insiemi sono enti reali, dato che i due insiemi rappresentati contengono diversi sottoinsiemi, devono essere due enti diversi. Dunque il 2 non può essere entrambe le cose: o è l'insieme descritto da von Neumann, o è l'insieme descritto da Zermelo. Quale delle due definizioni di numero ordinale dovrebbe essere quella vera? Il platonista non sa come rispondere: le due interpretazioni sono *isomorfe*, cioè descrivono lo stesso ordinamento e danno origine alla stessa aritmetica, quindi non può dire legittimamente che l'aritmetica derivata da una delle due interpretazioni sia falsa e l'altra vera, anche se *dovrebbe* dirlo (dato che i due sistemi descrivono enti diversi). L'antirealista, invece, può subito riconoscere che non c'è proprio alcun fatto in merito, nessuno dei due sistemi è vero in senso fattuale, e quindi entrambe le definizioni sono valide. Al che, l'unica alternativa logica è sposare una visione kantiana della matematica.

Il platonista, però, ha a disposizione varie risposte. Anzitutto, il problema evidenziato potrebbe essere esclusivamente epistemico, non ontologico: se gli enti matematici esistono, il 2 o è un ordinale di von Neumann, o è un ordinale di Zermelo (o altro equivalente), e anche se è vero che io non posso sapere quale dei due casi si dà, posso comunque credere che se ne dia uno. Questa inconoscibilità non è certo una tragedia: anche nelle scienze naturali ci sono teorie empiricamente indistinguibili, utilizzabili entrambe anche se solo una può essere quella fattuale. Inoltre, il platonista può anche negare l'esistenza degli insiemi. Per esempio, Zalta è contemporaneamente neo-logicista, platonista, e strutturalista: non crede nell'esistenza degli enti matematici come numeri ed insiemi (non esistono dato che non hanno proprietà intrinseche), ma crede nell'esistenza reale delle strutture matematiche, e numeri ed insiemi acquisiscono proprietà estrinseche in virtù del posto che occupano in tali strutture.

Il kantismo richiede di abbandonare la logica classica: l'argomentazione del miracolo

Alla fine, i problemi posti dalla filosofia matematica si riducono ad un confronto tra realisti e antirealisti. Riprendo allora l'esempio dell'ipotesi del continuo, che evidenzia un conflitto tra le due visioni ma anche tra filosofi e matematici.

L'ipotesi del continuo è vera o falsa? Per un realista, è una delle due cose; per un antirealista, non è né l'una né l'altra (per il momento). Una risposta è raggiungibile solo aggiungendo degli assiomi alla teoria degli insiemi. Ma quali assiomi? Al momento, nessuno di quelli proposti, indipendentemente dal loro accettare o rifiutare l'ipotesi, pare essere più ovvio o analitico dell'ipotesi del continuo stessa (in un caso del genere, tanto varrebbe porre ad assioma l'ipotesi). Immagino però che un giorno si scopra un assioma molto convincente che implica la verità dell'ipotesi: come dovrei considerarlo? Per i realisti, quell'assioma è sempre stato tale, solo che io

non lo sapevo; per gli antirealisti, invece, l'aggiunta dell'assioma ha di fatto creato la verità dell'ipotesi, ma nulla mi vietava di costruire un sistema diverso.

Ora, se la comunità scientifica stesse *costruendo* la matematica, si accorderebbe ad arbitrio sul da farsi, delibererebbe se l'ipotesi debba essere vera o falsa per poi porre un assioma di conseguenza. Ma la comunità non agisce così: la matematica non è soggettiva, non è un'opinione, uno non pone gli assiomi che vuole in base alla sua fenomenologia personale, al contrario si aspetta di *scoprire* la verità o falsità dell'ipotesi, di *scoprire* l'assioma corretto da aggiungere. Insomma, i matematici, contrariamente a molti filosofi della matematica, non considerano i vari formali sistemi possibili come equivalenti, non pensano che io possa dimostrare l'ipotesi del continuo *scegliendo* arbitrariamente un assioma tra tanti ugualmente possibili.

Passo ad un altro elemento di contrasto: l'infinito, attuale e potenziale. In linea di principio, un antirealista potrebbe credere nella sua esistenza: Cantor per esempio, pur essendo un formalista, credeva in un Dio che è attualmente e assolutamente infinito. Di norma però gli antirealisti, kantiani inclusi, hanno problemi ad accettare l'esistenza dell'infinito: se la matematica è solo una costruzione della mente, allora c'è un problema perché comunque questa mente è finita, non può contenere infinite idee; né basta che io *abbia idea* dell'infinito, perché la matematica dice proprio che ci sono infiniti insiemi, non che c'è l'*idea* dell'infinità degli insiemi. Brouwer, coerentemente con questa visione, nega l'esistenza di un infinito attuale, e a rigore avrebbe fatto bene a negare pure quello potenziale: se il segno ∞ rappresenta una quantità *potenzialmente infinita* di enti, come minimo devo riconoscere l'esistenza di infiniti *enti possibili*, che è cosa ben strana sia dal punto di vista ontologico (esiste ciò che *può* esistere?), sia dal punto di vista fisico (neanche l'universo può contenere infinite informazioni, deve avere un numero finito di stati possibili). Insomma, pare che solo un realista (o, eventualmente, un teista) possa accettare l'infinito a cuor leggero.

Un ultimo esempio di contrasto: le definizioni *impredicative*. Una definizione è impredicativa se caratterizza l'ente definito tramite un insieme che contiene l'ente medesimo: ripetendo un esempio di Russell, definire «l'inglese tipico» come «colui che è dotato di tutte e sole le qualità possedute dalla maggioranza degli inglesi» è impredicativo in quanto l'inglese tipico è incluso nell'insieme degli inglesi. Russell disapprovava questo tipo di definizione in quanto riconosceva in esse l'origine del suo paradosso; è però noto tramite Gödel e Ramsey che le definizioni impredicative sono perfettamente accettabili quando l'ente definito ha esistenza indipendente dalla definizione data: se esiste un inglese tipico, comunque costui esisterà in modo indipendente dall'insieme degli inglesi (è evidente anche perché potrebbe essere tutt'altro, potrebbe cambiare cittadinanza); al contrario, «l'insieme di tutti gli insiemi che non appartengono a sé stessi» non esiste indipendentemente dagli insiemi che non appartengono a sé stessi, né può essere altro che questo. Ma, in effetti, in matematica esistono già frotte di definizioni impredicative come questa. È un problema? Per l'antirealista sì, e bisogna fare qualcosa per evitare che sorgano paradossi; ma il realista non ha di questi problemi, perché per lui gli enti matematici sono enti esistenti in modo fattuale, e quindi indipendenti da qualsiasi definizione.

Riassumendo: mentre i realisti praticano la matematica così come si è sempre fatto, gli antirealisti chiedono che si usi un metodo diverso; mentre i realisti mantengono la logica e la semantica normalmente utilizzate moltiplicando il numero degli enti nell'universo, gli antirealisti mantengono un'ontologia snella e naturalistica cambiando gli assiomi logici e la teoria della verità associata. Oggi, io credo, questa seconda via sarebbe quella da favorire, bisognerebbe accettare che diversi campi del discorso rispondono a diversi principi logici e a diversi tipi di verità, per questo il pluralismo risulta sempre più attraente. Epperò, considerando la logica non revisionabile

razionalmente, mi sento più disposto ad accettare la via barocca e leibniziana, quella della moltiplicazione degli enti.

Vale la pena allora dare un'occhiata ad un'ultima argomentazione, che Putnam crea in favore del realismo scientifico, ma che può essere usata anche per difendere il realismo matematico dagli attacchi degli antirealisti.

Come detto, per Putnam *sarebbe un miracolo* se l'evoluzione naturale avesse permesso all'uomo di sviluppare credenze false, perché è ovvio che una credenza vera permette di sopravvivere meglio, ma sono ben poche le credenze false funzionali alla sopravvivenza. Generalizzando, *sarebbe un miracolo* se la scienza descrivesse il mondo in modo per lo più errato, perché è ovvio che avendo a disposizione una descrizione corretta del mondo io posso predire il comportamento futuro di un elettrone, mentre sono ben pochi i ritratti del mondo per lo più falsi che comunque portano a previsioni sistematicamente corrette. Al che, posso proporre lo stesso ragionamento per la matematica:

- P1** l'umanità ha sempre praticato la matematica *come se si desse il realismo*;
- P2** la matematica è una scienza formale di enorme utilità;
- P3** io posso derivare da premesse false un'infinità di metodi *errati* per praticare la matematica, e quasi nessuno di essi può essere utile alle scienze naturali;
- C** quindi *sarebbe un miracolo* se la matematica fosse stata praticata finora seguendo un metodo errato, quindi il metodo derivato dalla premessa del realismo è probabilmente basato su delle verità.

L'argomentazione del miracolo richiama quella d'indispensabilità: per assurdo, il naturalismo stesso richiede l'esistenza reale di enti matematici astratti. Infatti, se la scienza è in grado di descrivere la natura, allora la matematica è parte di questa natura; ma se la matematica non è parte della natura, allora devo cambiare il modo in cui la pratico, devo cambiare certi teoremi, alcuni dei quali potrebbero essere utilizzati nella scienza, e così anche il ritratto che la scienza dà del mondo deve cambiare, qualsiasi ritratto possa produrre la scienza con la matematica attualmente in uso non può essere quello corretto. Insomma, se cambio le conclusioni alle quali posso giungere in matematica, cambierò anche le conclusioni che raggiungo nella scienza.

Proprio come l'argomentazione d'indispensabilità, anche quella del miracolo si basa su una abduzione: dal successo di questo *platonismo metodologico* abduco la verità del platonismo stesso, semplicemente perché è raro che un metodo basato su falsità sia di successo. Nessuno però può garantirmi che non si dia quest'ultimo caso: in fondo, gli scienziati del passato parlavano del flogisto e dell'etere *come se esistessero*, e dandoli per veri avevano anche prodotto delle buone teorie, eppure nessuno oggi crede a certe cose. Quindi l'argomentazione non è cogente.

Certo, si può obiettare che io ho prove contro l'esistenza del flogisto, che non ho prove contro l'esistenza degli enti matematici, e che se non avessi prove contro il flogisto sarebbe legittimo credervi. Epperò proprio l'argomentazione del miracolo può essere rigirata per giustificare il rifiuto dell'abduzione. Infatti, è vero che solo un piccolo numero di insiemi, tra tutti quelli contenenti credenze false, può portare a comportamenti utili; è però vero anche che, tra tutti gli insiemi che portano a comportamenti utili, solo un insieme sarà quello composto da tutte e sole le credenze vere, tutti gli altri saranno o parziali o falsi. Di conseguenza, tra tutti i metodi di pratica matematica utili alla scienza, una fetta notevole avrà anche premesse false, cioè l'utilità del platonismo metodologico non rende affatto probabile la verità del platonismo ontologico.

Come nota Quine, insomma, le proposizioni della matematica non sono affatto più certe delle altre: se non riesco a immaginare come $2 + 2 = 4$ possa essere falso, è solo perché la matematica occupa una posizione centrale nelle mie convinzioni sul mondo, e non perché abbia un qualche statuto particolare rispetto alla fisica o alla teologia.

Quindi bisogna favorire teorie alternative

Alla fine, sia la prospettiva realista, sia quella antirealista, risultano indesiderabili. È necessario superare le metafisiche elementari che fanno da base a queste posizioni per riuscire a mettere insieme i loro aspetti migliori: poter affermare l'inesistenza degli enti matematici, e poter continuare ad usare la logica classica con una teoria della verità corrispondentista, il tutto senza passare per forme di proiettivismo o quasi-realismo come invece fa Blackburn in etica. Insomma, serve un *antirealismo conservativo*.

Arrivare ad una teoria di questo tipo, arrivare per esempio ad un intuizionismo che finisca per rivelarsi del tutto equivalente alla matematica tradizionale, non è di per sé impossibile, vi sono già state proposte in questa direzione (cfr. Nakano 2020). Ma qual è la metafisica che deve accompagnare una simile teoria? Se gli enti matematici non sono né entità astratte realmente esistenti, né pure e semplici costruzioni mentali, cosa dovrebbero essere? Diversi matematici e filosofi hanno proposto delle concettualizzazioni che sono ben diverse da entambi questi estremi.

L'astrattezza della matematica comunica un grande senso di libertà, e questo, una volta interiorizzato, può manifestarsi anche nel vagabondaggio e nell'eremitaggio. Ted Kaczynski, prodigio della matematica, decise di ritirarsi a vivere in una capanna in mezzo alla foresta nutrendosi degli animali che riusciva a catturare (poi, stimolato dalla lettura di Conrad, prese ad affermare il suo luddismo mandando pacchi bomba a destra e a manca). Grigori Perelman, in seguito alla risoluzione della congettura di Poincaré (un *problema del millennio*), avrebbe vinto la medaglia Fields, il premio del Clay Mathematics Institute, dell'European Mathematical Society, dell'International Congress of Mathematicians, premi in denaro per un valore di diversi milioni di dollari; ma lui ha sempre rifiutato tutto, ha rinunciato ad una vita da milionario per mantenere la libertà di starsene in una casa popolare con l'anziana madre ad occuparsi di cose religiose. Secondo una leggenda popolare, il fisico scomparso Ettore Majorana avrebbe scelto liberamente di vivere come un senzatetto, standosene sulla strada a offrire aiuto ai bambini coi compiti di matematica. Ecco infine Paul Erdős, *l'uomo che amava soltanto la matematica*: avendo rinunciato come San Francesco ad ogni bene terreno, vagava di casa in casa, da un amico matematico all'altro, offrendosi ad ogni tappa di aiutare nella dimostrazione di qualche teorema, e finendo così per diventare uno dei matematici più prolifici di sempre, secondo solo ad Eulero. E proprio Erdős, a sua stessa insaputa, offre con questo suo stile di vita una delle più originali concezioni degli enti matematici.

Erdős è un matematico sociale: celibe a vita, senza fissa dimora, privo di interessi al di fuori del suo lavoro, con all'attivo 1500 articoli pubblicati insieme a più di cinquecento co-autori, per lui la matematica è inseparabile dal rapporto sociale esistente tra i matematici. Anche se parlava come un platonista, immaginando una sorta di libro metafisico contenente tutte le dimostrazioni più eleganti, il suo modo di praticare la matematica fa pensare invece a un'arte sociale, fa pensare ad un linguaggio. Posso immaginare allora una *matematica del processo* nello stile di Whitehead: la matematica non è una sostanza, è invece definita da un processo esteso nel tempo, è definita dal

suo stesso cambiamento e dalla sua evoluzione, la quale è a sua volta determinata dalle relazioni interne alla comunità dei matematici.

Anche Popper, benché non fosse matematico, propone una visione alternativa interessante. Secondo lui, le formule matematiche come « $2+2=4$ » esprimono non una, ma più proposizioni, e mentre alcune di queste si riferiscono a fatti oggettivi, altre rimandano a costruzioni sociali. Di conseguenza, il dibattito tra realisti e antirealisti è per lui frutto di un fraintendimento: entrambe le teorie sono corrette, solo che parlano di proposizioni diverse (anche se espresse dallo stesso insieme di simboli). Il problema di questa visione è che il linguaggio matematico non pare affatto ambiguo, è anzi l'ideale per evitare l'ambiguità. Chiedo: Wiles ha dimostrato l'ultimo teorema di Fermat? Secondo la teoria popperiana potrei dire di no in quanto l'interpretazione del teorema potrebbe essere cambiata nel corso dei secoli, oppure potrei dire che ha dimostrato solo una delle interpretazioni del teorema. È invece ovvio che Wiles ha dimostrato l'unico ultimo teorema di Fermat esistente, non ci sono più teoremi nascosti negli stessi simboli.

Putnam ha proposto invece uno *strutturalismo modale* per gli enti matematici: le strutture che li contengono sono possibili, non attuali, cioè quando dico «3 è un numero primo» intendo dire qualcosa come «se esistesse il numero 3, sarebbe primo» (e si noti che, nel mondo possibile in cui esiste, il 3 potrebbe essere anche un ente fisico anziché astratto). Questa interpretazione non va però intesa come un tipo di logicismo: non c'è la volontà di ridurre la matematica alla logica modale, e comunque, considerando che per Quine la logica modale è «insiemistica sotto mentite spoglie» (forse perché certi assiomi modali riflettono certe teorie insiemistiche), una riduzione alla logica modale potrebbe non essere una riduzione in primo luogo.

Cantor, infine, sosteneva un aristotelismo dal gusto tomista: per lui la matematica è una creazione della mente umana, ma comunque tutto ciò che il matematico può scoprire esiste già nella mente di Dio (infatti interpretava le sue stesse scoperte come ispirate). Questo (ma pure la teoria vista con Erdős) permette di discutere la matematica tramite fatti sociali o psicologici senza darle un'esistenza oggettiva in natura.

...

È vero che i testi di interesse storico sono spesso molto utili per impadronirsi degli strumenti intellettuali della materia: si pensi agli "Elementi" di Euclide, che rimangono il modo migliore per imparare a pensare la geometria; . Credo però che affrontare questi testi in modo diretto non sia utile, vi sono altri modi per imparare la logica. Un testo di interesse storico ma più accessibile dei Principia e utile per comprendere il programma di Frege è "Introduzione alla Filosofia Matematica" di Russell, un testo da leggere piano prendendo molti appunti. Infine, non è una cattiva idea provare anche a studiare direttamente della matematica seria: si può partire con "Cos'è la Matematica" di Courant e Robbins per poi arrivare a impegnarsi su manuali come "Analisi Matematica" di Enrico Giusti e "Geometria Analitica e Algebra Lineare" di Anichini e Conti.

La variante più estrema del platonismo è il matematicismo, la filosofia che vede nella matematica il fondamento della realtà. Questa idea, da Pitagora e Platone, passando per

Cartesio e Leibniz, arriva ad un pensatore moderno come Alain Badiou (L'Essere e l'Evento). Il matematicista più ardito è Max Tegmark, che in "L'Universo Matematico" afferma direttamente che la matematica costituisce la natura basica dell'universo e che tutte le strutture matematiche sono ugualmente esistenti (si veda anche "The End of Time" di Julian Barbour). Penelope Maddy è una platonista più moderata ("Realism in Mathematics"); e similmente lo era Paul Erdos, del quale vi è una bella biografia, "L'Uomo che Amava solo i Numeri". Si veda anche il classico "Apologia di un Matematico", libro di G. H. Hardy molto amato.

L'argomentazione di Benacerraf contro il platonismo si trova nel suo paper "What Numbers Could Not Be"; l'argomentazione potente di Field si trova invece in "Science Without Numbers". Di contro, l'argomentazione di indispensabilità di Putnam (e Quine) si trova in "Matematica, Materia e Metodo", mentre la risposta ad Hardy sulle cicale è in "Are There Genuine Mathematical Explanations of Physical Phenomena?" di Baker. Sul tema del potere predittivo della matematica ha scritto anche Wigner in "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"; a lui ha risposto Hamming in "The Unreasonable Effectiveness of Mathematics". In merito al fatto che gli scienziati stessi non abbiano impegno ontologico verso gli enti inosservabili si vedano, quali esempi di antirealismo, "Scientific Representation", "The Empirical Stance" e "The Scientific Image" di van Fraassen.

Passando all'aristotelismo, il libro più importante della tradizione logicista è "Principia Mathematica" di Russell e Whitehead, da affiancare a "The Nature and Meaning of Numbers" di Dedekind e alla "Ideografia" e a "I Fondamenti dell'Aritmetica" di Frege; questa tradizione è oggi ripresa di Crispin Wright, soprattutto in "Frege's Conception of Numbers as Objects". Il curatore della Stanford Encyclopedia of Philosophy, Edward Zalta, è un logicista molto particolare, a cavallo col platonismo, e molto influenzato dai presocratici e da pensatori stravaganti come Meinong (si veda il suo "Abstract Objects"); lo stesso valga per Mayberry ("The Foundations of Mathematics in the Theory of Sets"). Il più importante formalista è invece David Hilbert: per il suo progetto si veda, per esempio, "The Foundations of Geometry", e vari suoi saggi quali "The Grounding of Elementary Number Theory", "The Logical Foundations of Mathematics", "The New Grounding of Mathematics" contenuti in "From Kant to Hilbert" di William Ewald. Anche Carnap ("Foundations of Logic and Mathematics") era un formalista, e così Haskell Curry ("Outlines of a Formalist Philosophy of Mathematics").

I primi passi verso un recupero delle posizioni kantiane sono stati fatti (involontariamente) da Poincaré in "Science and Hypothesis". Il più importante kantiano moderno, cioè intuizionista, è Brouwer, i cui lavori sono raccolti nei "Collected Works", ma se ne trovano molti anche in "From Frege to Godel: Source Book in Mathematical Logic" di Heijenoort; ma per uno scritto più amichevole si veda anche "Vita, Arte e Mistica". Suo erede è Heyting ("Intuitionism. An Introduction"). Tra i finitisti, altra visione kantiana, ricordo van Bendegem ("Finite, Empirical Mathematics, Outline of a Model"); tra i costruttivisti, invece, ricordo Goodstein ("Constructive Formalism - Essays on the Foundations of Mathematics"), Paul Ernest ("Social Constructivism as a Philosophy of Mathematics"), Lorenzen ("Constructive Philosophy") e Martin-Lof ("Philosophical Implications of Type Theory").

Per un'analisi dal punto di vista psicologico si veda "Da Dove Viene la Matematica" di Lakoff. Il principio di Hume si trova nell'ampilissimo "Trattato sulla Natura Umana". Peano ha esposto i suoi assiomi in "Arithmetices Principia". La più bella edizione degli Elementi di Euclide (ma solo per i primi sei libri) è quella di Oliver Byrne, che si trova per esempio nella collana Bibliotheca Universalis della Taschen (cioè tra i libri d'arte): Byrne, con sapienza didattica, presenta i teoremi di Euclide con illustrazioni che riescono a combinare in modo sorprendente la chiarezza scientifica con la magnificenza estetica; inoltre, per favorire l'intuizione

geometrica, dove Euclide scrive "l'angolo a...", "il lato c...", Byrne sostituisce le parole in mezzo al testo direttamente col disegno.

Il paradosso di Quine si trova in "Paradox". Hofstadter espone un'ipotesi sugli effetti del teorema d'incompletezza sulla mente nei libri "Anelli nell'Io" e "Godel, Escher, Bach": sono libri di terza cultura, espongono tesi originali di interesse accademico ma sono scritti in modo divulgativo e chiunque può fruirne.

4. 4. Filosofia della mente

Cos'è la mente

Frege è il nonno della filosofia analitica, abbiamo detto con Dummett: l'uso della logica per riflettere sulla matematica sta alla radice di tutta la filosofia analitica. La riflessione sui fondamenti della matematica, però, porta subito la riflessione all'attività della mente: anche Husserl arriva alla fenomenologia, all'analisi dei fenomeni mentali, riflettendo sullo stesso argomento. Questa connessione tra filosofia della matematica e filosofia della mente può trovarsi già in Kant: per Kant la matematica è speciale rispetto a tutte le altre discipline per la sua natura di conoscenza "sintetica a priori", l'unica altra disciplina a condividere questa particolare tipologia è la metafisica, e la filosofia della mente è la corona della metafisica. Sia Frege che Husserl sono anti-psicologisti, non credono affatto che la matematica possa essere spiegata dalla psicologia: quando gli analitici parlano della mente lo fanno in modo molto diverso dagli psicologi, per gli analitici i pensieri sono attitudini proposizionali, il cedere che qui ci sia un tavolo è l'aver l'attitudine della credenza nei confronti della proposizione "qui c'è un tavolo", ed è per questo che per gli analitici non esiste mente a prescindere dal linguaggio. Quest'idea, però, non è sempre stata comune: Mill, per esempio, pure riflette contemporaneamente su matematica e mente, e la sua conclusione è effettivamente che i fondamenti della matematica siano di tipo psicologico; oggi si ritorna a prendere ispirazione da questi pensatori, e si torna anche in ambito analitico a parlare di mente a prescindere dalle attitudini proposizionali e quindi dal linguaggio.

Questo richiede un chiarimento, perché in questo campo, che è oggi di gran lunga il più complesso e il più discusso di tutta la filosofia analitica, il fraintendimento infesta un sacco di ricerche: cos'è, esattamente, la mente in filosofia?

Vi sono due diversi modo di intendere questo termine.

Nel primo modo, la mente può essere intesa tramite i suoi processi: cioè la mente è l'insieme delle credenze, dei desideri, delle sensazioni, delle decisioni, e quant'altro normalmente avviene nella mente. Quando si intende la mente in questo modo bisogna tenere in considerazione due punti. In primo luogo, si tratta di un tipo di mente che è generalmente osservabile: se io desidero un gelato posso comunicarlo, e anche se non lo comunico il fatto che io vada a prenderne uno e lo divori in pochi secondi dimostra che lo desidero. In secondo luogo, è un tipo di mente che è indipendente da qualsiasi particolare struttura biologica, ed è quindi una mente totalmente distinta dall'attività neurale o processi analoghi: io posso programmare un computer tramite degli algoritmi tali che, analizzando certi dati, scelga una certa risposta tra più opzioni possibili, e così si può dire che il computer stia "decidendo" e manifesti quindi in questo senso una mente, ma posso farlo anche con una pluralità di supporti diversi, che arrivano a manifestare una mente tramite diversi percorsi. Questa concezione della mente è legata al "comportamentismo" per il quale appunto ciò che possiamo discutere della mente è manifesto nel comportamento, ed è quindi pubblicamente conoscibile.

Nel secondo modo, la mente può essere intesa come, per definizione, ciò che è soggettivo, ciò che è accessibile solo al soggetto. È ovvio che ognuno di noi ha una certa prospettiva su mondo: io e l'amico siamo davanti ad un quadro, lui sta alla mia destra, e io so allora che la mia esperienza del quadro è tale da avere un angolo leggermente diverso rispetto alla sua esperienza, so che se mi mettessi al suo posto potrei avere una esperienza più simile a quella che ha lui, infatti so che io vedo la luce che entra nei miei occhi e lui la luce che entra nei suoi, insomma so che lui ha delle esperienze coscienti anche se, per definizione, nessuno se non lui potrà mai esperirle, e io non potrò mai esperire se non le mie proprie esperienze coscienti. Queste esperienze coscienti vengono chiamate "qualia" e penso che tutti ne abbiamo una

definizione intuitiva: se io guardo la mia mano, io ho l'esperienza della mano che però non è la mano in sé, io so che la mia mano è fatta di atomi e cellule che però io non vedo affatto, ho un'esperienza della mano che costituisce il fenomeno (da cui il termine fenomenologia), e non posso vedere il noumeno che sta alla base di tutte le esperienze. Non dobbiamo fraintendere pensando che i qualia siano semplicemente un modo di spiegare la percezione, come se dicessimo che la luce che entra nel mio occhio crea un nuovo oggetto astratto, e che i qualia che esperisco davanti a un quadro sono una cosa, i miei ricordi tutt'altra, e le mie attitudini proposizionali (come "la credenza" verso una certa proposizione) altra ancora: i qualia sono né più né meno che quelli che i filosofi dei secoli passati hanno sempre chiamato "fenomeni", se io sono cosciente dei miei ricordi allora quei ricordi sono qualia, se sono cosciente dei miei pensieri allora quelle mie attitudini proposizionali sono qualia, negare che i qualia esistono equivale a negare che gli uomini hanno esperienze coscienti.

Bisogna fare attenzione anche all'uso della parola "coscienza". Ovviamente c'è un senso morale di coscienza, il "senso morale", e questo non ci interessa. La coscienza può essere intesa come "uno stato di veglia" e, come "autocoscienza", come la capacità di "metacognizione", di analizzare la propria mente stessa. Questo modo di intendere la coscienza è associato al primo modo di intendere la mente, è infatti una funzione che potenzialmente si può replicare anche su un computer come quelli che usiamo normalmente; non dobbiamo però confonderla con la coscienza intesa come la capacità di avere qualia, che è il senso associato al secondo modo di intendere la mente.

Noi umani abbiamo senza dubbio una mente in entrambi i sensi. Un computer potrebbe avere una mente nel primo senso, ma non ci pare proprio che abbia una mente anche nel secondo senso, manifesta comportamenti ma non crediamo che abbia esperienze soggettive. Magari ci sono oggetti che hanno una mente nel secondo senso, ma non ce ne accorgiamo perché non ne hanno una anche nel primo senso: in effetti, il primo senso è necessario affinché ci si possa rendere conto dell'esistenza della mente nel secondo senso, ma uno può ben avere esperienze soggettive anche se queste non si manifestano in alcun comportamento.

In ciò che segue, quando non lo specifico, intendo la mente nel secondo senso, in senso fenomenologico, non cognitivo.

Bisogna sottolineare che, decidendo di partire da questa definizione di mente, rinuncio anche a dare una panoramica complessiva del settore della filosofia della mente, che è enorme e per essere affrontato adeguatamente richiederebbe da solo diversi volumi. Infatti, una gran fetta di ciò che si discute all'interno della filosofia della mente riguarda la mente dal punto di vista cognitivo, coloro che parlano dell'identità della mente col cervello lo fanno quasi sempre discutendo solo l'aspetto cognitivo della mente; similmente, quando filosofi come Dennett dichiarano di non credere all'esistenza di qualia lo fanno quasi sempre intendendo i qualia come proprietà intrinseche non-fisiche e non-rappresentazionali, e non come fenomeni.

Leggere ricerche di filosofia della mente può essere molto faticoso, frustrante e noioso anche perché raramente i filosofi specificano esattamente a quale interpretazione di questi termini si stanno riferendo: danno per implicito che si capisca dal contesto, ma non sempre è così, anche perché ci possono notare quasi dei "dialetti" nel linguaggio della filosofia della mente, il modo in cui i filosofi dell'università di Oxford usano i termini (che vengono dalla lunga e gloriosa tradizione della filosofia del linguaggio ordinario), per esempio, non è il modo in cui li usano i filosofi australiani (molto influenzati da David Lewis e David Armstrong, filosofi materialisti ma altamente originali), che a loro volta parlano in modo diverso dai filosofi di Harvard (che è la terra di Putnam e Quine). Il modo in cui parlo della mente in queste pagine è più simile a quello dei pragmatisti americani in quanto l'introduzione del termine "qualia" (non dico "quale"

al singolare perché è ambiguo) inteso come fenomeno si deve a Peirce; siamo però anche ai confini della filosofia analitica, in sovrapposizione con la fenomenologia.

Il problema mente-corpo e il caso di Phineas Gage

Il problema che sta al cuore di tutta la filosofia della mente è questo: qual è la relazione tra il corpo e la mente.

Noi sappiamo, da un'amplissima serie di dati, che la nostra mente superviene sul nostro cervello, cioè sappiamo che ad un cambiamento nella struttura del cervello o nella sua attività corrisponde un qualche cambiamento nella nostra esperienza mentale, e che non è possibile un cambiamento nella nostra esperienza mentale senza una certa attività cerebrale.

Sono moltissimi gli esempi che ci fanno rendere conto di questo fatto: un ictus, la malattia di Alzheimer, la febbre alta, il bere troppi caffè, sono tutte cose che ci mostrano come ad un cambiamento neurale corrisponda un cambiamento mentale. Un metodo a priori per rendersi conto della necessità della supervenienza può essere questo: è chiaro che la nostra esperienza ha un legame causale con il mondo esterno, cioè se io vedo una sedia è perché questa sedia causa la mia esperienza; la sedia fisica, però, non può interagire causalmente con una mente astratta (se potesse avere interazione causale non sarebbe astratta), ciò che può fare è riflettere la luce in modo tale da colpire il mio occhio e causare una serie di eventi neurali nel mio cervello, quindi solo se ho questi eventi neurali posso avere anche un'esperienza mentale perché gli oggetti fisici possono interagire solo col mio corpo, la mente deve essere mediata dal corpo (il che non esclude comunque la possibilità delle allucinazioni, vi sono anche esperienze senza rapporto causale, quelle che non sono esperienze "di qualcosa").

Storicamente, il caso più importante e sconvolgente, quello che più ha contribuito a creare una coscienza della supervenienza della mente sul corpo, è stato quello di Phineas Gage.

Operaio, il 13 settembre 1848, un mercoledì, alle quattro e mezza del pomeriggio, stava inserendo una carica esplosiva in una roccia che bloccava il percorso della linea ferroviaria in costruzione quando questa, spinta e compattata con un'asta di metallo, esplose sparandogli l'asta lunga più di un metro e pesante sei chili proprio sotto lo zigomo sinistro, l'asta trapassò il cranio da parte a parte distruggendogli il loro frontale sinistro. Incredibilmente, dopo pochi minuti dall'incidente, era cosciente e in grado di camminare e anche chiacchierare, venne portato alla vicina città e assistito subito da Edward Williams, medico e magnate ferroviario. Attraverso la ferita, stando a quanto riporta Williams, era possibile vedere il cervello pulsare; quando Gage vomitò la pressione fece uscire dalla ferita mezza tazza di materia cerebrale. Alle sei del pomeriggio fu portato da Harlow, il quale pulì la ferita rimuovendo coaguli, frammenti di cranio, e circa trecento grammi di cervello. Inizialmente lucido, le sue condizioni iniziarono a peggiorare il giorno successivo e nel giro di due settimane era pressoché moribondo.

Nonostante la ferita fosse costantemente drenata si formò un enorme ascesso cerebrale, il pus continuava ad uscire dalla testa e gli amici avevano già preparato bara e vesti idonee alla sepoltura. Harlow, che era tra i pochi medici dell'epoca ad avere preparazione in merito ad accessi cerebrali, lo operò di nuovo, asportò altro tessuto cerebrale marcescente e fece uscire un'intera tazza di pus. Dieci giorni dopo la nuova operazione Gage era di nuovo in grado di alzarsi; dopo alcune settimane fu giudicato in via di guarigione e fatto tornare a casa. Un anno dopo l'incidente aveva ancora una certa sensazione di debolezza, ma riusciva a lavorare normalmente; nel 1860 iniziò ad avere convulsioni frequenti che lo portarono alla morte all'età di 36 anni. Nel 1866 Harlow chiese alla famiglia di riesumare il corpo per poterne avere il teschio, ritenuto di interesse medico, e la famiglia accettò subitaneamente (parenti serpenti); il teschio e l'asta sono tutt'ora esposti al Warren Museum.

Poco dopo il ritorno a casa, parenti e amici descrivono una serie di cambiamenti in Gage. Aveva una maggior difficoltà a valutare le dimensioni e il denaro, una memoria minore, difficoltà di concentrazione, una diminuita abilità sociale. Di particolare interesse, per l'epoca, erano i cambiamenti relativi alla personalità: dopo l'incidente Gage sarebbe diventato di colpo volgare, irrispettoso, pigro, impaziente, impulsivo, disinibito, un uomo totalmente diverso dal Gage furbo e intraprendente di prima. Questo mostra l'impatto che può avere sul comportamento un danno cerebrale.

Quanto siano accurate queste testimonianze, comunque, è dubbio: la frenologia, la pseudoscienza che cercava correlazioni tra la forma del cranio e la personalità, era rigogliosa all'epoca e molti cercavano a forza delle conferme nel caso; altri credevano invece che i danni cerebrali non avessero alcun effetto sulle più alte funzioni cognitive, e questo poteva portarli al contrario a sminuire a forza il cambiamento. Il cambiamento c'è stato, ma è probabile che sia stato momentaneo, ridotto ai primi anni dopo l'incidente.

È incredibile il genere di danni ai quali il cervello può adattarsi. È vero, a volte basta un po' di sangue nella zona sbagliata, un danno minuscolo in qualche piccola struttura profonda, per causare all'istante la cessazione di funzioni vitali; eppure il cervello riesce a riadattarsi anche danni enormi, anche dopo la distruzione di metà dei suoi tessuti riesce a trasformarsi e a garantire la possibilità di ritornare ad avere una vita normale. Il cervello non è come il chip di un computer, strutturato in modo fisso: le sue cellule si trasformano di continuo, creano di continuo nuove connessioni o ne perdono, è fatto per apprendere e così, anche se un danno cerebrale distrugge certe abilità, comunque questa "plasticità cerebrale" permette in molti casi di ristrutturare le cellule restanti per sopperire al danno.

Un altro caso incredibile, che vediamo più sinteticamente, è quello di Kim Peek. Nato con pesanti deformità al cervello (il corpo calloso che connette i due emisferi era totalmente assente), questa sua condizione si era manifestata con profonde disabilità, tanto che era incapace di abbottonarsi la camicia da solo; eppure, tanto grandi erano le disabilità, tanto e più erano i doni eccezionali che aveva ricevuto: leggeva libri interi in poche ore, era in grado di leggere con un occhio la pagina a destra e con l'altro la pagina a sinistra contemporaneamente, e ricordava in modo letterale pressoché tutto ciò che leggeva; aveva letto più di 12.000 libri imparando a memoria un mare di nozioni sulla storia del cinema, sul mormonismo e la Bibbia, sulla Storia e la letteratura, conosceva a memoria le piantine di varie città strada per strada nonché il codice postale di pressoché tutte le città negli Stati Uniti; sapeva suonare il pianoforte a istinto, e ricordava nota per nota canzoni anche ascoltate una sola volta anni prima; aveva capacità di calcolo prodigiose, gli si poteva dire una qualsiasi data storica e lui era in grado di calcolarne istantaneamente il giorno della settimana.

Quello di Kim Peek è un caso di savantismo: una condizione nella quale grandi disabilità e grandi abilità si accompagnano. Anche nei casi di emisferectomia si assiste, pur molto raramente, a casi del genere: pare che il cervello, dovendo recuperare una certa funzionalità riadattando altri circuiti prima adibiti ad altro, può creare delle connessioni insolite che portano ad un funzionamento stravagante ed eccezionale. Tommy McHugh, umile costruttore, spingendo troppo mentre era in bagno (stiamo attenti, eh), ha finito per causarsi una doppia emorragia cerebrale, l'aumento della pressione gli ha fatto scoppiare un'arteria nel lobo frontale e anche in quello temporale; rimasto in coma per un mese, quando si è risvegliato è stato colto da un pervasivo interesse per l'arte e la poesia, che ha praticato con successo fino alla sua morte. Orlando Serrell venne colpito alla testa da una palla da baseball; ignorando il mal di testa che lo ha tormentato dopo la botta, come questo scomparve Serrell si rese conto di aver di colpo acquisito la capacità di calcolare il giorno della settimana di qualsiasi data e di ricordare in modo estremamente dettagliato tutto ciò che ha fatto giorno per giorno anche decenni prima.

Derek Amato stava tuffandosi in piscina non accorgendosi dell'acqua troppo bassa, ha sbattuto la testa causandosi concussioni e perdita dell'udito; poco dopo, però, si è reso conto di aver sviluppato una forma di sinestesia, una strana connessione tra le esperienze sensoriali che può portare a vedere immagini e colori ascoltando suoni o sentire sapori guardando quadri, e di essere in grado di suonare il piano molto meglio di quanto fosse mai riuscito a fare prima, tanto da iniziare a comporre musica. Anthony Cicoria, chirurgo, è uno dei pochi ad essere stato colpito da un fulmine ed essere sopravvissuto per raccontarlo; salvato da un'infermiera, recuperò velocemente tutte le sue funzionalità, solo che ora, in più era travolto da un'incontrollabile desiderio di suonare il piano, cosa che iniziò a fare subito da totale autodidatta e con notevole successo. Jason Padgett, lavoratore quarantenne, stava uscendo dal karaoke quando venne pestato di brutto da dei rapinatori, la stessa notte suo padre morì di cancro e poco dopo suo fratello scomparve senza lasciare notizie di sé; lo shock gli fece sviluppare ossessioni e fobie, ma, contemporaneamente, si rese conto che, guardandosi la mano, riusciva a vedere intricati pattern geometrici che non riusciva a spiegare a parole; iniziò a studiare matematica all'università, e finalmente trovò un modo per esprimere ciò che vedeva, rendendosi conto tra l'altro di riuscire a disegnare frattali a mano grazie alle superiori abilità di visualizzazione sviluppate.

Tutti questi sono casi estremamente rari: è come esporsi a radiazioni e diventare Mr. Fantastic invece di sviluppare qualche tumore. Epperò sono anche casi che, oltre a farci rendere conto del collegamento esistente tra mente e corpo, riempiono di stupore e speranza: perché, anche nel caso di danni cerebrali, la probabilità che questo non sia solo una disgrazia, ma anche un dono, è bassa ma non nulla.

Zombie filosofici

La supervenienza delle mente sul cervello, anche se convincente, non è ovviamente sufficiente a farci identificare la mente col cervello: da qui nasce il problema mente-corpo.

Leibniz fa il famosissimo esempio di un mulino: io posso camminare tra gli ingranaggi in un mulino, e non vedo altro che questi; similmente, se io ingrandissi a dismisura un cervello potrei vedere l'attività dei neuroni, vedrei il segnale elettrico che passa da uno all'altro e le sostanze chimiche galleggiano tra le sinapsi, ma in nulla di tutto ciò vedrei la mente, da nessuna parte potrei vedere l'esperienza soggettiva associata a questi eventi elettrochimici. Può essere che il mio dolore sia il riflesso di certi gruppi di neuroni che "sparano" segnali in sincronia; questo evento fisico, però, in nulla assomiglia alla mia esperienza soggettiva del dolore, ad uno e all'altro assegno proprietà totalmente diverse, una cosa è totalmente oggettiva e l'altra è totalmente soggettiva una è un evento fisico e l'altra una esperienza...

Sembra, insomma, che l'esperienza mentale sia separabile da qualsiasi evento fisico, sembra cioè che, anche se l'evento mentale si accompagna sempre ad un certo evento fisico, cionondimeno gli stessi identici eventi fisici potrebbero avvenire senza essere accompagnati da alcuna esperienza soggettiva. Insomma: è possibile che esista un essere umano esattamente come me, con un cervello identico al mio, nel quale avvengono gli stessi eventi, e quindi un soggetto che dice ciò che dico io, fa ciò che faccio io, scrive quel che scrivo io, solo che, a differenza di me, non ha alcuna esperienza cosciente di quello che fa, è uno zombie.

Questo esempio, piuttosto istintivo nella sua semplicità, è stato sviluppato dal più influente filosofo della mente contemporaneo, David Chalmers (laureato in matematica, ma specializzatosi in neuroscienze sotto Hofstadter), che un tempo sembrava una rockstar (e in effetti lo era) ma che da quando si è tagliato capelli e barba come Sansone ha perso un sacco di carisma sia in quanto ad aspetto che a stile filosofico.

Chalmers si appoggia ad una teoria semantica sviluppata da David Kellogg Lewis.

Putnam e Kripke credono che il fatto che l'acqua sia H₂O sia una necessità metafisica a posteriori, credono che l'essere H₂O sia una caratteristica essenziale dell'acqua. Una obiezione sollevata da Searle era che in realtà se uno non sa che l'acqua è H₂O quando dice "acqua" fa un'affermazione vaga, solo dopo aver fatto la scoperta può decidere se l'essere H₂O è parte essenziale del significato oppure no. David Lewis cerca di mettere insieme entrambe le intuizioni: ad un primo livello, quando dico "acqua" l'intensione è il suo senso generico, è "roba liquida, trasparente, insapore"; ad un secondo livello, invece, quando dico "acqua" sono aperto ad accettare nel senso di acqua tutto ciò che scoprirò essere tale indagando la sostanza che così chiamo; quindi, ad un primo livello posso dire che "l'acqua sarebbe potuta non essere H₂O", ma ad un secondo livello non posso.

Chalmers fa un discorso simile in merito alla mente. Ci è evidente che questi "zombie filosofici" ci sono concepibili: infatti li stiamo concependo. Allora, sono possibili nel senso in cui è possibile che l'acqua non sia H₂O, è possibile in quanto ad intensione di primo livello. Al che, o è possibile anche come intensione di secondo livello, e allora non c'è alcun collegamento necessario tra corpo e mente, oppure c'è un collegamento necessario e quindi si può dire che corpo e mente formino un tutt'uno.

La maggior parte dei filosofi riconosce che gli zombie filosofici sono una possibilità logica, ma nega che siano una possibilità metafisica, come che l'acqua non sia H₂O. Adesso il problema sta nel spiegare perché dovrebbe essere così, perché sono metafisicamente impossibili, cosa c'è che impedisce l'esistenza di esseri del genere: non pare un compito facile (da notare, comunque, che nella semantica di Kripke, a differenza di quella di Lewis, in casi fenomenologici come questi la concepibilità è già possibilità, quindi non ci sarebbe questa distinzione).

Un esempio simile è quello dei qualia invertiti. Cioè, anziché immaginare zombi senza qualia, possiamo immaginare che, ad un certo stato cerebrale, corrispondano più esperienze soggettive possibili; possiamo immaginare cioè che vi siano persone che, pur avendo un cervello che funziona in modo identico al nostro, hanno esperienze coscienti molto diverse dalle nostre, mentre noi vediamo le mele rosse rosse, loro le vedono come verdi, ma non possiamo rendercene conto perché chiamano "rossa" l'esperienza che noi chiameremmo verde. Anche in questo caso, se esseri del genere sono concepibili, allora capiamo come i qualia sono nettamente distinguibili dagli stati cerebrali.

Mente, corpo e pipistrelli

Se tra mente e corpo c'è un collegamento necessario, in cosa consiste esattamente questo collegamento? Viene da dire che il cervello in qualche modo genera la mente. Questa risposta non è affatto chiara, e bisogna distinguere due sue possibili interpretazioni.

Da un primo punto di vista, si può intendere la risposta come relativa al fatto che la coscienza soggettiva correla con qualche struttura cerebrale, cioè che io ho esperienze coscienti solo quando il mio cervello è strutturato in un certo modo. Questa è una banalità piuttosto evidente: quando io dormo non ho alcuna esperienza cosciente, quando sono sveglio ne ho (a meno che non sia sonnambulo e quindi non-proprio-sveglio), osservando il mio cervello quando dormo e facendo un confronto con quando sono sveglio è possibile individuare dei tipi d'attività molto diversi, e si può affermare allora che la coscienza dipende da alcune di queste differenze.

Questo, però, è un discorso che rimane in superficie: ciò che si fa, in questo caso, non è altro che individuare quegli eventi neurali sui quali la coscienza superviene, ma non ci spiega affatto, dal secondo punto di vista, in che modo l'attività neurale genera l'esperienza cosciente; è una spiegazione di tipo cognitivo, ma non riesce ad affrontare il problema da un punto di vista fenomenologico. Cioè, anche se il concetto di causa rimane vago, ci è intuitivamente

chiaro come un calcio causa il volo del pallone, ma come un certo evento neurale dovrebbe causare un'esperienza soggettiva non è affatto chiaro.

Ciò che dobbiamo chiederci, seguendo Nagel, è com'è essere qualcosa: io posso conoscere tutto del funzionamento del cervello di un pipistrello, ma questo non mi è sufficiente a capire "com'è" il poter sentire gli ultrasuoni, sapere quale evento neurale correla con l'esperienza uditiva di un ultrasuono non mi dice nulla su com'è l'avere questa esperienza in prima persona come qualia. Hofstadter è uno dei pochi uomini ai quali non piace il cioccolato: lui non capisce com'è mangiare il cioccolato sentendolo come buono, forse sente un sapore diverso dal mio perché ha una lingua strutturata in modo diverso, forse sente lo stesso ma lo percepisce come cattivo, ma anch'io vorrei capire com'è il sentire buono il pesce, non riesco a capirlo.

Questa domanda è forse la più complessa di tutta la filosofia, tanto che viene chiamata direttamente "il problema difficile": l'uomo ci riflette da millenni, abbiamo avuto enormi sviluppi nelle nostre conoscenze scientifiche, eppure il problema non si è mosso di un millimetro, in tutta la nostra esplorazione del mondo non abbiamo trovato neppure un minuscolo indizio di come un evento elettrochimico oggettivo possa causare l'esistenza di una prospettiva soggettiva.

Negli ultimi anni sono stati in molti, in effetti, a pensare che il "problema difficile" non può essere risolto, il tipo di risposta che potrebbe ricevere è assolutamente irraggiungibile da un essere umano; c'è un modo nel quale il corpo crea la mente, ma il poterlo descrivere e capire è lontano dalle nostre capacità tanto quanto il risolvere un integrale è lontano dalle capacità di una mucca. Tra questi "nuovi misterianisti" vi sono Chomsky, che è il più grande linguista di tutti i tempi oltre che filosofo e teorico politico; Friedrich Hayek, forse il secondo o terzo più grande economista del secolo scorso; Edward Witten, che è invece il maggior fisico vivente; e Steven Pinker, uno dei più famosi psicologi viventi e magnifico scrittore.

Molti però non si sono arresi, i tentativi di risposta sono stati innumerevoli, e innumerevoli sono le scuole di pensiero in merito. Per praticità, e rimanendo al livello più generale possibile, possiamo individuare quattro possibilità.

Per i fisicalisti esiste solo il corpo, e la mente è generata da esso in qualche modo.

Per i dualisti corpo e mente sono due sostanze indipendenti, connesse in qualche modo ma autonome.

Per i monisti corpo e mente sono due aspetti di una stessa sostanza dotata delle proprietà di entrambi.

Per gli idealisti, infine, esiste solo la mente, e il corpo è illusorio.

Fisicalismo riduzionista

Un tempo vi erano filosofi materialisti, filosofi che credevano che tutto ciò che esiste fosse la materia. Questa posizione, con lo sviluppo delle nostre conoscenze scientifiche, si è dimostrata sempre più assurda e insostenibile: la materia non esiste in modo necessario, la materia scompare e ricompare casualmente migliaia di volte al secondo in base al variare della distribuzione dell'energia nello spazio vuoto, materia ed energia sono appunto due facce della stessa cosa, e non ci potrebbe neanche essere la materia se non accettassimo l'esistenza di alcune altre cose, delle proprietà, della radiazione, dei campi, dello spaziotempo, delle leggi fisiche... La posizione del materialismo, così, si è evoluta accettando queste precisazioni, è diventata una posizione più raffinata, quella del fisicalismo, cioè l'idea che ciò che esiste è ciò che è fisico.

Come nel caso del "naturalismo", anche "fisicalismo" è un termine vago: cosa vuol dire che esiste "ciò che è fisico"? Se è "fisico" ciò che è descritto dalle nostre teorie fisiche, be', la teoria è semplicemente falsa, le nostre teorie sono in continua evoluzione e sappiamo che ciò che

siamo giustificati a credere oggi non saremo più giustificati a crederlo domani; se invece è "fisico" ciò che la fisica arriverà a scoprire alla fine della sua storia, la teoria è così vaga che la si può usare per sostenere di tutto, non abbiamo alcuna idea di come sarà la fisica tra un miliardo di anni.

Sulla definizione di fisicalismo esiste una discussione infinita, tanto che si potrebbe dire che il tema più discusso all'interno del fisicalismo è proprio quale sia la sua definizione. È probabilmente meglio, per ora, accontentarci di una definizione intuitiva: diciamo che ciò che esiste è l'insieme di quegli enti che sono contemplabili dalla fisica, gli altri enti descritti dalle altre materie (molecole, cellule, stati psicologici, istituzioni sociali...) sono semplicemente composti dagli enti fisici. L'idea del fisicalismo, insomma, è che se un mondo possibile è identico al nostro in tutti i fatti fisici, e non si postulano altri fatti allo stesso livello o ad un livello più basso di quelli fisici, allora quel mondo è identico al nostro in tutti i fatti.

Possiamo dire che per il fisicalismo gli enti di maggiore complessità "emergono" dal fisico. Cosa vuol dire che "emergono"? Prendiamo un quadro di Seurat: esso è null'altro che un insieme di puntini colorati, se osservato da lontano però ci sembra di vedere un'immagine significativa; possiamo dire, così, che l'immagine emerge dai puntini. Una parola scritta è solo un insieme di lettere, la parola emerge dalle lettere. Più scientifico: vi sono delle proprietà che io riconosco al livello chimico, riconosco il pH per esempio, la misura di acidità e basicità; io so però che, sia pure con enorme sforzo e scrivendo descrizioni di insopportabile complessità, io potrei descrivere il pH usando solo enti fisici, potrei descrivere la distribuzione degli ioni di idrogeno all'interno della soluzione. La riduzione, allora, vuol dire questo: che un certo fatto può essere descritto totalmente tramite i termini di un'altra classe di fatti; quindi, per esempio, un fatto chimico è riducibile ad un fatto fisico se ogni evento chimico è identico ad un certo insieme di eventi fisici.

Il fisicalismo non è quindi solo una filosofia della mente, è una visione del mondo complessa, un'ontologia che risponde alla domanda "cosa esiste?" al livello più generale possibile. Proprio per questo, però, ha un riflesso immediato in filosofia della mente. Quando interpretato in modo fisicalista tale posizione ci dice che tutto ciò che è mente (desideri, conoscenze, e anche le esperienze coscienti) possono essere descritte totalmente in termini fisici, la mente può essere descritta totalmente tramite il corpo, e una volta fatto ciò non esiste proprio nient'altro da descrivere, e quindi la mente non è nulla più che il corpo stesso concettualizzato in diverso modo, esattamente come una statua è null'altro che il suo marmo solo interpretato selezionando diverse proprietà.

Questa interpretazione riduzionista del fisicalismo ha diversi e ovvi problemi. Ciò che abbiamo detto sui qualia, l'esempio del pipistrello di Nagel, ci mostrano che i qualia non sembrano essere affatto descrivibili totalmente in termini fisici, una descrizione completa del cervello non pare essere in grado di dirci tutto ciò che ci è necessario sapere sulle esperienze soggettive, deve esserci "qualcosa in più", nei qualia, oltre al corpo.

Un esempio, prima sostenuto e poi rifiutato da Frank Jackson, è quello della "stanza di Mary": Mary è una scienziata che è nata e cresciuta all'interno di una stanza in cui ogni cosa è solo bianca e nera, nulla è colorato e Mary non ha mai avuto contatto con nulla di colorato proveniente dal mondo esterno alla sua stanza; ciò nonostante, è stata nutrita, cresciuta, e la si è fatta studiare tutto ciò che l'umanità conosce in merito ai colori, è una grande scienziata su questo argomento e conosce tutta la fisica relativa alle lunghezze d'onda, tutta la biologia e la neurologia relativamente al funzionamento dell'occhio e all'analisi delle informazioni provenienti da esso, eccetera. La domanda è: quando Mary uscirà per la prima volta dalla stanza, e avrà per la prima volta esperienza soggettiva del colore, imparerà qualcosa in più, o non avrà invece accesso a nulla in più di quello che aveva già imparato dai suoi studi?

Le risposte variano. David Lewis afferma che Mary non apprende nuovi fatti, un "sapere che", apprende invece un "sapere come", un come ricordare e identificare i colori per esempio. Un'altra strategia sta nel distinguere i fatti conosciuti dal modo in cui li si conosce: se io imparo a memoria una parte dell'elenco telefonico divento un esperto sugli abitanti di Conegliano, e posso dire di sapere che Alessio abita a Conegliano; se però io non mi ricordo di essere Alessio, non posso dire di sapere anche che "io" abito a Conegliano; quando mi ricorderò chi sono, apprendereò anche che "io abito a Conegliano", ma questo non rappresenta un fatto in più rispetto ad "Alessio abita a Conegliano", è solo un diverso modo di esporre lo stesso fatto (similmente, il vedere rosso è solo un diverso modo di conoscere un certo evento neurale che Mary conosceva già). Il problema con queste risposte è che io posso conoscere che "io abito a Conegliano" solo tramite la conoscenza di quello che effettivamente è un nuovo fatto (che io sono Alessio); similmente, è difficile spiegare come sia possibile arrivare a conoscere un nuovo "sapere come" senza conoscere anche un nuovo "sapere che". Ognuno può fare un esperimento su di sé: come appare una sfumatura media tra il violetto e l'indaco? Dire "sfumatura media tra violetto e indaco" e immaginare questo colore sono chiaramente cose diverse, vedo che mi danno accesso a proprietà diverse, può essere che siano semplicemente due modi di rappresentare un certo evento nel cervello (come la statua e il suo marmo sono due modi di rappresentare un certo oggetto) ma non possono essere due modi di rappresentare una identica proprietà di quell'evento, devono essere comunque proprietà diverse: i qualia sembrano rappresentare davvero una conoscenza in più rispetto al mero funzionamento del corpo, i qualia soggettivi sono qualcosa di diverso dagli eventi neurale oggettivi, la soggettività dei qualia è essenziale come è essenziale l'oggettività per gli eventi neurali, e quindi uno non può essere ridotto all'altro.

Non solo i qualia, ma anche l'intenzionalità è un problema per questo tipo di fisicalismo. Com'è possibile che la nostra mente, se è null'altro che una serie di eventi fisici, si "riferisca" a qualcosa? Perché il mio "pensare a Mozart" è diretto a qualcosa, mentre il creare un programma che richiama un'immagine del mio cane Mozart è solo una simulazione senza alcun potere di riferimento? In generale, i fisicalisti sono attratti dalla teoria disposizionalista: il significato di una parola è l'essere disposti ad usarla in un certo modo, il significato della parola "bianco" sta nell'essere disposti ad usarla davanti ad oggetti bianchi; similmente, il riferimento del mio pensiero "Mozart" sta nell'essere disposto ad usarlo quando percepisco il mio cane. Vi è però un problema con questa teoria, e cioè che il nostro uso delle parole può anche essere errato. Io posso vedere un oggetto bianco tramite un vetro rosso, e io, ingannato, posso dire "questo oggetto è rosso": se il significato della parola sta nel modo in cui sono disposto ad usarla, be', pare proprio che io sia disposto ad usare la parola "rosso" anche davanti ad oggetti bianchi; ma chiaramente non è così, la mia disposizione non è sufficiente a rendere merito del significato delle parole, il disposizionalismo non può essere una teoria completa; ma quindi il fisicalista non può neanche accettare che la nostra intenzionalità stia nelle nostre disposizioni, e diventa per lui molto più difficile trovare un modo per giustificarla.

Il fisicalismo come l'abbiamo definito, però, rimane tale fintanto che continua ad affermare che ciò che esiste è solo l'insieme degli enti fisici: l'idea che tutti gli altri enti siano automaticamente riducibili al fisico, che siano totalmente descrivibili usando solo termini fisici, non è necessaria. Com'è possibile? Se non esiste null'altro che il fisico, non dovrei anche poter spiegare tutto il non-fisico (il chimico, il biologico, il psicologico...) tramite il fisico? Non necessariamente.

Fisicalismo non-riduzionista ed emergenza

Vi sono due tipi di emergenza. Quando è possibile effettivamente ridurre un campo all'altro si parla di emergenza debole: per esempio, la chimica emerge dalla fisica, ma si tratta di una emergenza debole perché, in linea teorica, ogni affermazione della chimica può essere riscritta come affermazioni fisiche, sia pure tremendamente lunghe e complesse; vi sono composti chimici le cui molecole si dispongono in modo tale da far emergere un quadrato perfetto, oppure ne fiocchi di neve le molecole si distribuiscono in modo tale da creare complesse forme simmetriche, ma avendo una adeguata conoscenza delle leggi fisiche si può arrivare a spiegare perché le molecole hanno caratteristiche tali da distribuirsi in quel modo anche senza alcuna guida particolare. Sia chiaro: le proprietà chimiche emergenti esistono, hanno poteri causali, e sono chiaramente distinte dalle proprietà fisiche; però gli eventi fisici sono giustificati solo da cause fisiche (non chimiche), e le proprietà chimiche emergenti sono causate in ogni istante da proprietà fisiche.

Alcuni tipi di emergenza debole possono essere osservati anche creando semplici programmi col proprio PC. L'informatico Chris Langton immagina una "formica" che si muove secondo regole molto semplici: per esempio, può partire da un pixel del proprio computer, se il pixel davanti è bianco la formica lo colora di nero e poi gira a destra, se ne trova uno nero lo colora di bianco e gira a sinistra, e continua a muoversi secondo queste due regole all'infinito. Seguendo regole così semplici le formiche creano a volte strutture sorprendentemente complesse: a volte la formica, dopo aver girato in modo caotico, entra in loop e crea una "autostrada" infinita; altre volte crea strutture perfettamente simmetriche, altre ancora immagini che si auto-incorniciano, o figure il cui interno viene riempito, o che si espandono a forma di frattale, o di spirale, comunque facendo emergere pattern molto complessi visibili solo nell'insieme.

[inserire alcuni esempi]

Il matematico Stephen Wolfram è rimasto stregato da questi "automi cellulari", programmi che basandosi su pochissime regole mostrano l'emergenza di comportamenti complessi; ha scoperto che alcune regole molto semplici possono creare dei pattern osservabili in natura, e pensa che pressoché tutta la natura possa emergere in modo simile, crede che partendo da una manciata di regole in tutto sia possibile far emergere un intero universo con una fisica proprio come la nostra, immagina così un "nuovo tipo di scienza" (che è anche il titolo del suo librone, un manifesto lungo mille pagine) che, studiando questi automi cellulari, può arrivare a capire che posto occupa il nostro universo tra tutte le possibilità computazionali, un "attacco a forza bruta", come un hacker può cercare di azzeccare una password inserendo tutte le password possibili, così uno scienziato può cercare di scoprire le leggi dell'universo calcolando i possibili insiemi di regole da cui può emergere un sistema fisico come quello studiato. Wolfram non chiede affatto di cambiare il modo in cui facciamo scienza, ma il suo approccio può ricordare la "fisica digitale" immaginata da uno degli uomini più strani che siano mai vissuti, Frederick Parker-Rhodes, che era linguista poliglotta (leggeva in più di venti lingue), micologo, mistico quacchero, ma soprattutto un matematico che, con una certa sorpresa, è riuscito a derivare costanti relative alle quattro forze fondamentali della nostra fisica da un semplice algoritmo. Tutti questi, comunque, sono esempi di emergenza debole: conosciamo le regole che fanno muovere la formica di Langton, e quindi possiamo calcolare questi fenomeni emergenti a partire dalla base, è possibile mostrare nella maggior parte dei casi come da quelle regole basiche si può implicare la regola del pattern che osserviamo al livello macroscopico. Questo, però, non è necessariamente vero in ogni caso, perché esiste anche la possibilità di una

emergenza forte, proprietà che emergono da altre senza però che sia possibile ridurle o capire come le regole basiche hanno portato alla comparsa di quelle proprietà.

Il matematico John Horton Conway, morto da poco, è il creatore del "gioco della vita" (esiste anche la variante "gioco della vita senza morte", che purtroppo non era quello che stava giocando lui). Si tratta di una serie di algoritmi che, partendo da regole molto semplici come gli automi cellulari, mostrano un comportamento emergente che in qualche modo appare "vivo": vi sono oggetti che si muovono, che si scontrano, che creano copie di sé stessi, e quant'altro. Ora, noi possiamo descrivere facilmente il comportamento complessivo di un "gioco della vita", e conosciamo anche le semplici regole che lo governano. Dovremmo allora essere facilmente in grado di capire se il programma prevederà mai un certo stato futuro: per esempio, se io so dalla descrizione del suo comportamento macroscopico che il programma crea un pattern a spirale, dovrei essere in grado, conoscendo semplicemente le regole che governano il gioco, di calcolare se quel pattern apparirà prima o poi. In realtà, non è così, è un problema indecidibile: pur conoscendo tutte le regole che governano il gioco non posso calcolare se un certo pattern comparirà mai anche se questo è invece evidente quando provo a descrivere il gioco a livello macroscopico, posso sapere se un certo pattern apparirà mai solo facendo girare il programma e osservandolo finché non compare.

Si può obiettare, giustamente, che si tratta di un esempio puramente matematico, non c'è motivo per cui una cosa del genere debba avere un corrispettivo in fisica. Orbene, uno degli studi più interessanti degli ultimi anni su questo tema è quello di Mile Gu, il quale ha dimostrato che è possibile costruire sistemi fisici il cui comportamento complessivo, se fosse computabile a partire dalle proprietà delle parti, allora certi problemi computazionali che noi sappiamo essere indecidibili sarebbero invece indecidibili. Diciamolo in modo diverso. Io posso immaginare di creare un certo sistema fisico, posso immaginare di distribuire in un certo spazio certe particelle, o certe molecole, con certe proprietà che io conosco tutte. Orbene, l'interazione tra queste particelle o molecole è tale da far emergere un comportamento complessivo particolare, un po' come la distribuzione secondo pattern simmetrici emerge nel sistema complessivo dall'interazione delle molecole d'acqua nei fiocchi di neve (le quali di per sé non hanno alcun pattern particolare). A differenza del caso dei fiocchi di neve, però, in questi sistemi fisici le proprietà che emergono non sono calcolabili a partire dalle parti: cioè, mentre nei fiocchi di neve, conoscendo tutte le proprietà delle loro molecole, posso potenzialmente calcolare che si distribuiranno in quel modo, in quest'altro caso anche conoscendo tutte le proprietà dei componenti del sistema fisico io comunque non potrei prevedere queste proprietà emergenti.

Questo di per sé non dimostra l'esistenza di una emergenza forte (quella che abbiamo descritto in entrambi i casi è ancora emergenza debole) però ce la rende comprensibile: vediamo cioè che non c'è proprio nulla di logicamente aberrante nell'idea che certe proprietà emergenti possano essere totalmente inspiegabili tramite le proprietà delle parti, cioè che non ci sia alcun algoritmo che spieghi come dalle proprietà delle parti si giunga a quelle proprietà emergenti. Ciò che distingue l'emergenza forte dalla debole è che l'insieme manifesta dei poteri causali che le sue parti intese come un semplice ammasso non potrebbero manifestare, e che, pur accettando la mente come non-basica (cioè composta da enti non-mentali), la riconosce comunque come fondamentale (non riducibile ad enti non-mentali, ma indipendente e "nuova"); e vediamo che in effetti è possibile che dei sistemi fisici manifestino proprietà che non possono essere descritte tramite i termini che descrivono il comportamento delle loro parti, ed è allora possibile che anche il sistema del nostro cervello sia tale da far emergere proprietà mentali anche se le sue componenti non hanno proprio nulla di mentale né è possibile calcolare come tali componenti creino le proprietà mentali.

Questo tipo di emergenza è molto preoccupante per filosofi e scienziati. Se le proprietà emergono in senso forte, allora non solo supervengono in modo irriducibile, ma gli si può anche assegnare potere causale, cioè il fatto che un sistema si comporti un certo modo viene giustificato non semplicemente dalle proprietà delle sue parti, ma viene giustificato in modo diretto dalle proprietà emergenti. Noi non sappiamo se nel nostro universo esistono realmente sistemi del genere; se però esistono, allora è quasi magia, sono proprietà che emergono in un modo che ci è davvero incomprensibile, emergono come il genio emerge dallo sfregamento della lampada di Aladino.

Ciò nonostante, il fisicalista ha tutto l'interesse ad accettare una interpretazione di questo tipo della sua posizione: solo in questo modo, infatti, riesce a spiegare contemporaneamente perché la mente superviene sul corpo (perché emerge dal fisico e non c'è altro) ma il mentale rimane perfettamente distinguibile e irriducibile al fisico (perché è una emergenza forte); cioè, vi è una differenza epistemologica tra mente e corpo, ma non ontologica, noi concettualizziamo mente e corpo in modo diverso anche se si tratta di due interpretazioni dello stesso identico fatto, e questo fatto è puramente e semplicemente fisico. Anche se può sembrare "magia" non c'è proprio nulla di anti-scientifico in questo, il concetto è stato profondamente analizzato negli ultimi decenni e i filosofi sono giunti alla conclusione che l'emergenza forte in nulla contraddice i concetti che stanno alla base della scienza, al contrario la Storia della scienza stessa ci mostra come già in passato sia stato normale stipulare nuove leggi che spiegassero proprietà non riducibili ad altre già note, né è strano che proprietà emergenti incarnino leggi in grado di interferire con il comportamento di livelli più basilari della realtà.

In più, l'emergenza forte può rendere merito di diverse ipotesi che abbiamo già visto. Com'è possibile il libero arbitrio? La nostra fisica non pare essere in grado di contenere alcun libero arbitrio, ma se il libero arbitrio è una proprietà che emerge in modo forte, allora non solo è possibile, ma possiamo spiegare anche com'è possibile esercitarlo, visto che il nostro comportamento non viene causato semplicemente dalle proprietà delle nostre molecole, ma, al contrario, parte del comportamento delle nostre molecole dipende dal libero arbitrio emergente, con una causazione in senso inverso (dal tutto alle parti). I teoremi di incompletezza di Godel possono già suggerirci qualcosa del genere: Hofstadter, pur non sostenendo questo tipo di argomentazioni, nota che, così come non è possibile dimostrare la consistenza dell'aritmetica di Peano all'interno della stessa ma è facile farlo usando un diverso sistema più generale, così, quando parliamo della mente, possono esserci dei concetti, come appunto quello di libero arbitrio, che possono essere spiegati ad un alto livello, cioè guardando l'insieme, ma che non possono ricevere alcuna spiegazione, neanche fosse infinitamente lunga, tramite i livelli più bassi come quello fisiologico. Volendo, infine, ci si può spingere anche al di là della metafisica, e usare questo tipo d'emergenza per accordare naturalismo e non-naturalismo metaetico: le proprietà etiche, così, possono emergere fortemente dalle naturali, in modo irriducibile, spiegando com'è possibile che esistano proprietà etiche non-naturali anche all'interno di una metafisica che non accetta l'esistenza di entità non-naturali.

Funzionalismo

Il fisicalismo è spesso associato alla teoria dell'identità tra mente e cervello. Questa teoria non può essere interpretata, ovviamente, in modo letterale: per la legge di Leibniz, dire che mente e cervello sono identici vuol dire che sono caratterizzati dalle stesse identiche proprietà, e questo è assurdo, la mente è soggettiva e il cervello è oggettivo, l'esperienza mentale della luce è un certo colore e non una lunghezza d'onda mentre da un punto di vista oggettivo ci sono lunghezze d'onda e nessun colore, le proprietà sono chiaramente diverse. Mente e cervello possono essere considerati identici in un altro senso: nel senso che ad un certo evento neurale

corrisponde sempre e solo un certo evento mentale in modo tale che uno sia logicamente equivalente all'altro. Per esempio, un fisikalista potrebbe sostenere che ogni volta che le fibre C "sparano" in sincronia uno prova dolore, e ogni volta che uno prova dolore le fibre C sparano in sincronia, non ci sono casi in cui si dà una cosa senza che si dia l'altra, e dire di una persona che "prova dolore" o che "le sue fibre C sparano in sincronia" è la stessa cosa.

Putnam individua una facile e ovvia obiezione a questa teoria. Gli umani hanno un cervello strutturato in un certo modo, anche se ogni uomo lo ha leggermente diverso; le varie specie di animali hanno tutti un cervello diverso uno dall'altro, e spesso è strutturato in modo molto diverso da com'è strutturato il cervello dell'uomo; eppure la nostra intuizione è che le esperienze mentali siano simili per tutti, crediamo che l'esperienza soggettiva del dolore sia simile per i vari uomini, e sia anche simile per gli animali, io sono convinto che ciò che io sento quando ho mal di pancia sia simile a quello che sente l'amico quando ha mal di pancia, e sono convinto che sia anche simile a ciò che il mio cane sente quando ha mal di pancia. Questo è il problema delle "realizzazioni multiple": un certo stato mentale, come il dolore, può corrispondere ad una gran varietà di stati neurali diversi, e non è vero quindi che lo stato mentale è perfettamente equivalente ad un certo stato cerebrale.

La soluzione di Putnam, allora, è quella di ragionare per funzioni, non per stati. Qual è la "funzione" del dolore? Proprio in senso matematico: una funzione è ciò che collega gli elementi di un certo insieme ad un certo altro, e il dolore allora è quella funzione che collega certi stimoli, come una martellata in testa, a certe reazioni, urla, pianto, e quant'altro di comportamentale colleghiamo al dolore. Cosa realizzi esattamente questa funzione, però, è irrilevante: può essere un cervello umano, un cervello di delfino con strutture molto diverse dalle mie, può essere un complessissimo chip di un qualche computer, può essere qualsiasi cosa, l'unica cosa importante è che il sistema incarni una certa funzione, cioè che sia in grado di recepire certi stimoli e di reagire a questi in modo appropriato.

È curioso che proprio Putnam, in diverse fasi della sua carriera, abbia proposto esperimenti mentali contrari a questa teoria. La stessa idea dei "cervelli in vasca", per esempio, è anti-funzionalista: i nostri cervelli e i cervelli in vasca svolgono le stesse funzioni, quindi non dovrebbe essere vero che le loro parole si riferiscono a fatti diversi, come invece sostiene; dice che un computer che mostra una certa parola sullo schermo non si sta riferendo ad alcunché, com'è intuitivo, ma se il chip di un computer stesse svolgendo la stessa funzione che svolge il mio cervello quando cerca di riferirsi a qualcosa allora non c'è motivo per cui debba essere così. Prendiamo un muro: ha così tante molecole disposte in modo non troppo ordinato che, a volerlo, sarà possibile trovare pattern equivalenti a qualsiasi programma, ma non per questo diremmo che il muro ha dei programmi, che calcola o pensa o rappresenta proposizioni; certi fanatici esamina in modo così ossessivo e approfondito foto sfocate della Sacra Sindone che alla fine ci vedono scritte di ogni tipo, ma anche fosse, queste non rappresenterebbero niente, non sarebbero proposizioni che si riferiscono a qualcosa, semplicemente è un oggetto composto da così tante fibre e così contaminato che ognuno a forza di incrociare gli occhi uno può arrivare a trovarci qualsiasi cosa cerchi. Alla fine, Putnam rifiuta il funzionalismo perché riconosce che è solo un complesso tipo di materialismo: infatti, se la mente è ciò che svolge una certa funzione, ed è ovvio che nel corpo umano ciò che gestisce il collegamento dalla martellata all'urlo è il cervello, allora dobbiamo affermare direttamente che la mente è uguale al cervello.

Un altro esempio che lui stesso fa contro il funzionalismo è quello dei qualia invertiti: supponiamo che una notte i collegamenti tra i neuroni nel mio cervello si scambino in modo tale che io ora veda il mondo in negativo, ma che mantenga la memoria dei colori positivi "così com'erano quando ero bambino": io ora posso allenarmi in modo tale che, colpito da una certa

lunghezza d'onda, quando io ho l'esperienza del verde, comunque affermi "vedo rosso"; in questo modo, ad un certo stimolo continua a seguire la stessa risposta, la funzione della mia esperienza potrebbe sembrare la stessa di prima, ma è ovvio che rimane qualcosa di diverso. Per dirlo in modo più specifico: la mia sensazione che prima aveva il ruolo funzionale di "segnalare il rosso oggettivo" (mi segnala gli oggetti rossi, è infatti la sensazione del rosso) ora ha il ruolo di "segnalare il verde oggettivo" (mi segnala gli oggetti verdi pur restando la sensazione del rosso); la mia teoria, però, è che l'evento mentale del "verde soggettivo" sia quell'evento che ha la funzione di "segnalare il verde oggettivo", quindi secondo tale teoria dovrei dire che la sensazione del verde è quell'esperienza che una volta legavo al rosso; ma questo è ovviamente sbagliato, e lo sarebbe anche se perdessi la memoria di come percepivo i colori una volta.

David Lewis fa invece l'esempio di un masochista: il dolore, in lui, occupa un diverso ruolo funzionale che in noi, noi quando sentiamo dolore urliamo e vogliamo liberarcene mentre lui non dice niente e lo desidera, questo però non è sufficiente a dire che il suo non sia dolore. Abbiamo già citato la stanza cinese di Searle: la stanza svolge nel suo complesso la funzione di una persona che sa il cinese, perché le vengono dati in pasto certi simboli e risponde a questi scrivendo cose appropriate, eppure non possiamo affermare che la stanza sappia il cinese perché la stanza non ha una mente!

Ci sono altri esperimenti mentali che puntano in questa direzione. Simile all'esperimento di Searle è la "blockhead" di Ned Block, simpatico filosofo che è stato proprio allievo di Putnam ad Harvard.

Ripetiamo che la mente è un qualcosa di soggettivo: non è mai accessibile se non in prima persona. La possibilità degli zombie filosofici discende da ciò: io non posso mai sapere in modo certo che le altre persone hanno una mente proprio come me, potrebbero anche non averla e io non avrei modo di saperlo, così come potrebbe essere che i sassi abbiano una mente ma io non lo credo erroneamente. Com'è possibile allora valutare se un computer, che pare agire in modo intelligente, abbia una mente, oppure no? Una risposta estremamente comune tra gli psicologi all'inizio del Novecento era: tramite il comportamento. Quest'idea era molto attraente, era fortemente sostenuta dal maggior psicologo del secolo scorso, Skinner, nonché dai due più grandi filosofi, Wittgenstein e Quine, e dal maggior informatico, Turing.

Ed è anche l'idea del maggior informatico del secolo scorso, Turing, il grande matematico che ha decifrato il sistema della macchina nazista Enigma e che, accusato di omosessualità, si è suicidato, come Biancaneve, addentando una mela avvelenata (da cui il simbolo della Apple). Sua è l'idea di questo test: creiamo un computer che è in grado di sostenere una conversazione con un essere umano, facciamo "chattare" degli esseri umani, a distanza, con computer e con altre persone, quando un uomo non è più in grado di distinguere quando sta parlando con un uomo e quando sta parlando con un computer, allora abbiamo tutto il diritto di dire che quel computer ha una mente proprio come un uomo, perché non possiamo più distinguerlo tramite nulla di osservabile, e se continuassimo a sostenere che non ha una mente, be', nulla potrebbe farci cambiare idea e non saremmo più davvero aperti a quell'ipotesi. Test di questo tipo, in effetti, vengono fatti di continuo, e siamo riusciti a produrre chatbot notevoli che sembrano avere anche un comportamento creativo.

Già von Neumann, nel suo ultimo libro abbozzato poco prima di morire, evidenziava chiare differenze tra una macchina e il cervello. Una macchina, per quanto possa essere potente, ha un numero finito di stati, mentre un essere umano pare poterne avere infiniti, per esempio può pensare infinite proposizioni diverse; gli stati di un uomo, inoltre, sono strutturati (per esempio, se io posso pensare che Alessio ama Giada, posso pensare anche che Giada ama Alessio, il poter pensare uno mi dà la facoltà di poter pensare anche l'altro), ma per una macchina

generalmente non è così. Al tempo di von Neumann non pareva proprio che la mente potesse funzionare come un computer. Col tempo, però, le cose sono cambiate: per esempio Fodor ha fuso il computazionalismo (l'idea appunto che la mente funzioni come un computer), il rappresentazionalismo (l'idea che gli stati mentali rappresentino oggetti, cioè siano intenzionali) e il formalismo (che elaborino simboli secondo regole, come visto in matematica) con l'idea risalente al francescano Occam dell'esistenza di un "linguaggio del pensiero", creando così una teoria computazionalista che non va incontro alle obiezioni viste (infatti i simboli del pensiero sono combinabili in modi infiniti, come possiamo fare con le parole, e dalla funzione $f(a, b)$ possiamo ricavare sia una frase che l'altra). Al che, se effettivamente possiamo trattare la mente come un computer, non c'è motivo per il quale non possiamo accettare il test di Turing, se un computer si comporta in modo indistinguibile da una mente allora ha una mente.

Ned Block, però, è molto scettico verso il valore di test di questo tipo. In fondo, in ogni linguaggio, posta una certa domanda, esiste un numero limitato di risposte sensate che si possono dare. Potremmo immaginare magari che esista una lingua estremamente semplice, come un pidgin (una di quelle quasi-lingue che si sviluppano quando due popolazioni che non hanno un linguaggio comune vivono in contatto per un certo periodo, e si sforzano di trovare un metodo per comunicare), dove ad una prima domanda come "come stai?" seguono solo un centinaio di risposte possibili adeguate; a ognuna di queste cento seguono una ventina di risposte adeguate, e via così, finché la conversazione non arriva ad un limite e finisce. Per quanto l'albero delle risposte possa essere molto grande, comunque in un caso del genere non pare impossibile programmare un computer in un modo tale che abbia in memoria un "libro" con tutte le risposte accettabili possibili, e che selezioni in modo automatico, forse addirittura casuale, una risposta possibile. In questo caso, però, non ce la sentiremmo affatto di dire che tale computer è "intelligente", non diremmo che ha alcuna comprensione della lingua, nonostante ogni sua mossa sia perfetta, e svolga in tutto la funzione di un parlante competente. Si tratta di un caso simile a quello visto nell'argomentazione di Lewis per il libero arbitrio: può ben essere che un computer sia così raffinato da riuscire a mimare in tutto e per tutto il comportamento di un essere umano, e riesca quindi a collegare in modo adeguato ogni stimolo ad una risposta; epperò, ciò che costituisce effettivamente il suo funzionamento, il modo in cui davvero arriva a formulare queste risposte, è significativo, non sono disposto ad accettare che un comportamento intelligente che emerge però da un processo irrazionale sia sufficiente a definire il computer che così agisce come intelligente. Un computer che si limita ad eseguire un programma, che ad ogni domanda fa seguire una certa risposta memorizzata agisce in modo sintattico, manipola simboli, ma non ha semantica, non possiamo dire che sappia quello che vuole dire una certa parola, proprio come una formica che per caso ci pare disegnare un'immagine sulla sabbia.

Questo ha effetti, forse anche indesiderabili, ad ampio spettro. Un programma come quello immaginato da Block può essere infinitamente potente, ma in virtù del suo funzionamento non lo possiamo riconoscere come cosciente. La stanza immaginata da Searle, similmente, è una macchina di Turing, cioè una macchina in grado di elaborare qualsiasi programma elaborabile da qualsiasi altra macchina, e la nostra intuizione è che non possa essere cosciente: ma se una macchina di Turing non può essere cosciente, allora nessuna altra macchina può esserlo. Dovremmo dire allora, più umilmente, che non possiamo sapere se la stanza di Searle ha una mente che conosce il cinese. Questo però ci suggerisce direttamente che, in modo puro e semplice, non possiamo in nessun caso sapere se una macchina ha una mente oppure no. Un altro esempio, sempre fatto da Block, è questo: un uomo si colpisce il dito col martello e grida "ahia!", un altro fa lo stesso e grida "d'oh!", uno stesso stimolo è collegato a due reazioni

diverse e quindi lo stato mentale che collega le due cose deve essere diverso per il funzionalismo; questo, però, è abbastanza assurdo, ci pare ovvio che si tratta dello stesso dolore in entrambi i casi: in questo caso il funzionalismo fa l'errore di farci tenere in considerazione differenze che sono sì funzionali ma chiaramente irrilevanti. Di contro, potrebbe anche farci perdere di vista differenze rilevanti ma non-funzionali: non c'è particolare motivo per cui, se io e l'amico siamo entrambi in uno stato di dolore (entrambi ci siamo dati una martellata, entrambi gridiamo), lui non possa più di me percepire anche la sensazione di un particolare odore, cioè non c'è motivo per cui ad una stessa struttura funzionale non possa corrispondere una esperienza diversa, ma concludere che, essendo uguale la struttura funzionale, allora è uguale anche lo stato mentale, non rende merito di questa differenza esperienziale rilevante.

Un ultimo esempio, ancora da Ned Block, è quello del "cervello cinese". Nel nostro cervello le funzioni sono svolte da neuroni che si passano segnali elettrochimici: una certa struttura neurale istanzia una certa funzione, è un modo di rappresentare un certo algoritmo tanto quanto i fori su un disco o un circuito su una lastra di silicio. Non c'è però limite ai mezzi coi quali possiamo creare un computer: si possono trovare online esempi molto divertenti, si possono creare cancelli logici usando le tessere del domino, a seconda di quale tessera si butta giù, si inserisce il valore "1" o "0", mentre le tessere sono strutturate in modo tale da dare un certo output a seconda dell'input scelto, di modo tale che diventa possibile creare così un algoritmo in grado di fare calcoli. Or dunque, noi possiamo immaginare allora di prendere tutti gli abitanti della Cina, disporli su un grande campo a forma di cervello, e istruirli per passarsi l'un l'altro certi messaggi a certe condizioni, insomma li si struttura in modo tale da svolgere una funzione desiderata. Il punto è: diremmo noi che questo insieme di persone, svolgendo nel suo insieme una certa funzione, è anche, nel complesso, una mente? Possiamo accettare che esistano menti composte da altre menti, e che menti esistano all'interno di un'altra giga-mente? Ovviamente no: i singoli uomini che compongono la struttura hanno una mente, ma la struttura complessiva non ha una mente, è un insieme di uomini e null'altro, un insieme di uomini non crea un nuovo super-individuo con una nuova mente solo perché strutturato in un certo modo. Ci sono numerose risposte possibili a questa argomentazione. I Churchland, Patricia e Paul, i più scienziati e nichilisti tra i filosofi viventi, hanno risposto, per esempio, che la mente emerge quando le funzioni vengono svolte con una certa ed elevata velocità, il problema del cervello cinese è che le funzioni avvengono lentamente e non alla velocità alla quale avvengono in un computer o in un cervello organico. Dennett critica invece certe argomentazioni chiamandole "intuition pump", argomentazioni fatte per guidare le intuizioni verso la conclusione desiderata anche quando sarebbe possibile esporla in modo tale da far giungere alla conclusione opposta. D'altro canto, però, la filosofia analitica è in parte anche un'esplorazione delle nostre intuizioni, in filosofia analitica è raro trovare un'argomentazione più forte a favore di una teoria del suo essere conforme al senso comune; in filosofia della mente, poi, questo è ancora più vero: crediamo che tutti gli esseri umani abbiano esperienze coscienti, che quando dormono non ne abbiano, che cani gatti mucche siano coscienti, che anche gli insetti lo siano ma molto meno, che i computer non lo siano affatto, e tutto questo lo crediamo perché sono tra le intuizioni più forti e basiche che abbiamo e, essendo l'esperienza del tutto soggettiva, non abbiamo altro mezzo migliore per indagare questo argomento.

Volendo, si potrebbe continuare all'infinito trovando controesempi nei quali la funzione di una struttura è chiaramente insufficiente a giustificare l'assegnazione di una mente: potrei fare un programmino semplicissimo in cui, alla pressione di un pulsante, il computer fa apparire un "aaah!" sullo schermo, potrei magari anche inserire un piccolo pistone sotto questo pulsante in modo tale che il computer sia in grado di sapere quanto fortemente spingo, e possa usarlo per

cercare di rialzare il pulsante e ridurre la pressione; il computer sembra soffrire tanto che ad una pressione forte grida e si sforza di farla cessare, ma sarebbe solo un gioco, non l'avrei certo reso cosciente del suo dolore.

Oppure, proviamo a parlare della memoria, del ricordo, da un punto di vista puramente funzionale: se non ci diamo limiti di semplicità, è memoria tutto ciò che permette di conservare una nozione e di riorfrircela in seguito, ma questa funzione può benissimo essere svolta anche fuori dal cervello, è una funzione che fa anche un pezzo di carta; in effetti c'è chi prende questa idea molto seriamente parlando di una "mente estesa", anche il taccuino su cui scrivo sarebbe effettivamente una parte della mia memoria. L'idea che la mente formi un tutt'uno col suo ambiente ha le sue ragioni (se io creassi una simulazione perfetta del mio cervello, ma questa fosse totalmente isolata dall'ambiente, be', non si comporterebbe per niente come me, il mio cervello ha l'esperienza di una lepre perché davanti a me sbuca una lepre, chi volesse simulare l'attività della mia mente per prevedere il mio comportamento dovrebbe non solo copiare il mio cervello ma anche tutto l'ambiente, simulare un solo cervello è già simulare un intero universo), ma la teoria della mente estesa rimane stravagante, controintuitiva, e difficilmente accettabile.

Vi è però, in tutto questo discorso, un notevole fraintendimento. In effetti, il funzionalismo non è un'alternativa al fisicalismo, e le risposte di Block non sono argomentazioni contro il funzionalismo. Il fisicalismo non è affatto obbligato ad accettare una relazione uno-a-un tra stati cerebrali e stati mentali, non c'è motivo per il quale non possa accettare che uno stesso stato mentale soggettivo possa emergere da più sistemi fisici (non può accettare invece il contrario, come nel caso dei qualia invertiti, perché anche l'emergenza dovrebbe essere ribaltata). Il funzionalismo parla della mente da un punto di vista comportamentale, parla della mente come si manifesta in ciò che è osservabile, non parla mai della mente come esperienza soggettiva, parla degli aspetti cognitivi e mai degli aspetti fenomenologici. Le argomentazioni di Block accusano appunto il funzionalismo di assegnare esperienze soggettive anche a sistemi che ovviamente non hanno una mente; ma il funzionalismo non cerca di fare questo, non assegna una mente a nessun sistema, e se lo fa, se crede che assegnare ad un ente una mente del primo tipo, cognitiva, sia logicamente equivalente all'assegnare una mente del secondo tipo, fenomenica, cade in un fraintendimento dei termini (dovrebbe sostenere, a questo punto, che il mio inconscio è il conscio di qualcun altro, visto che realizza delle funzioni ma io non ne ho esperienza). Il funzionalismo, si perdoni il gioco di parole, non funziona per i motivi individuati da Block solo se si pensa che stia parlando della mente in senso fenomenologico, ma se lo fa in modo cognitivo invece le argomentazioni di Block non lo scalfiscono affatto; d'altro canto, però, se lo fa in modo cognitivo, non ci interessa, visto che ci stiamo occupando della mente fenomenologica.

Epifenomenalismo

Vi è una facile critica al fisicalismo non-riduzionista sviluppata da Jaegwon Kim, uno dei filosofi che aveva più a cuore il rispetto per il rigore scientifico: l'idea che questo tipo di fisicalismo sovradetermini, cioè preveda per uno stesso fenomeno due cause contrastanti, una fisica e una mentale.

Accettiamo che il mentale emerga dal fisico, che lo stato che ora ha il mio corpo faccia emergere un mio certo stato mentale. Noi normalmente accettiamo che la mente abbia un potere causale: il mio urlare in seguito ad una scottatura è causato dal dolore (uno stato mentale), per esempio. D'altro canto, il mio dolore emerge da una serie di eventi fisici, da un certo modo di comportarsi delle mie cellule. Che senso ha, allora, che io dica che è il dolore a causare il mio urlo? Non dovrei dire che sono i miei eventi cellulari a causarlo? Si può dire:

be', entrambi causano l'urlo, sia gli eventi cellulari che il dolore. Ma questa è sovradeterminazione: così come non ha senso spiegare un fulmine sia con la nostra fisica che con la rabbia di Zeus, e dobbiamo scegliere solo una delle due spiegazioni, similmente avviene con il nostro comportamento, che non possiamo spiegare contemporaneamente con un fenomeno fisico e uno mentale.

Per rappresentarlo in modo schematico: possiamo immaginare un evento mentale M che sopravviene su un fatto fisico F; l'evento mentale M porta ad una nuova esperienza M*, la quale sopravviene su un nuovo fatto fisico F*; il che però è come dire che l'evento mentale M* è stato causato sia da M che da F, da qui la sovradeterminazione.

Se si accetta questa argomentazione, allora esistono due possibilità. La prima sta nel negare questo tipo di emergenza, ma se lo si fa ci si espone alle argomentazioni di zombie filosofici e simili. La seconda, allora, è l'accettare l'"epifenomenalismo", cioè la teoria che vi siano sì proprietà mentali (che emergono dal fisico), ma che queste siano causalmente inefficaci, abbiamo esperienze ma queste non fanno alcuna differenza; le esperienze mentali sono come la segatura che vola via lavorando il legno, è impossibile che non si produca ma di per sé non serve a niente. Può sembrare un'idea scienziata, ma in realtà la base dell'idea è nata in ambito religioso: Leibniz, per esempio, sosteneva il "parallelismo psicofisico" per il quale fatti fisici e fatti mentali evolvono per conto loro senza alcun rapporto di causa, e se ci sembra che la mente abbia effetti causali è solo per l'armonia prestabilita da Dio che fa sì che mente e psiche siano in sincronia; similmente, il teologo Malebranche sostenne l'"occasionalismo" per il quale ogni evento è solo un'occasione d'intervento per Dio, Dio vede che il mio corpo è ferito e mi fa provare dolore mentale, vede che voglio alzare la mano e fa muovere il mio corpo, ma io, con la mia mente (come anche col corpo) non posso causare alcunché.

Praticamente tutte le scuole di pensiero (fisicalisti, monisti, dualisti, idealisti) potrebbero adottare una versione dell'epifenomenalismo; di fatto, però, tutti cercano di evitarlo perché ha caratteristiche fortemente indesiderabile. In primo luogo, è controintuitivo: la nostra forte impressione è che il nostro dolore sia causa del nostro urlo, sarebbe ben strano dire che il dolore è solo qualcosa in più, e che se per qualche motivo avessimo provato piacere avremmo urlato lo stesso. In secondo luogo, mina la nostra capacità di conoscere le altre menti: se la mente altrui non ha alcun effetto sul comportamento, be', come faccio a dire che hanno una mente e com'è che questa è strutturata? In terzo luogo, è difficile accordare ciò con l'evoluzione: se la mente non ha alcun effetto sul comportamento, perché abbiamo una mente? Non sarebbe stato più semplice ed economico non svilupparla, visto che tanto ci saremmo comportati esattamente nello stesso modo? Come nota William James, se la mente non ha effetto sul comportamento, perché abbiamo l'intuizione che ad una ferita normalmente segue, in tutti gli organismi sufficientemente complessi, una sensazione dolorosa, invece che una sensazione a caso (come sarebbe legittimo se ogni sensazione fosse irrilevante al comportamento, e quindi la selezione non possa favorire nessuna rispetto alle altre)? Forse c'è una connessione necessaria tra certi comportamenti e certe sensazioni, magari una certa struttura non può non avere una mente: ma perché tale connessione dovrebbe essere necessaria, e non contingente?

In realtà, il fisicalista non è obbligato né a rinunciare all'emergenza, né a sostenere l'epifenomenalismo: l'idea che sovradetermini è infatti frutto di un fraintendimento...

Dalla teoria dell'identità al dualismo delle proprietà

Era comune, all'inizio della storia della filosofia analitica, la teoria dell'identità tra corpo e mente, il che era inteso non solo nel senso che esiste una sola sostanza, ma anche un solo set di

proprietà: le proprietà mentali non sono altro che proprietà fisiche, il dolore non è null'altro che un certo evento neurale; quindi, in sostanza, un fisicalismo del tipo riduzionista visto all'inizio. Kripke presenta un argomento potente contro l'identità di mente e corpo. Prendiamo il calore. Possiamo dire che il calore è uguale all'energia cinetica media di un gruppo di particelle (in realtà non è proprio esatto, ma soprassediamo), quindi è per me legittimo dire che un certo corpo è caldo anche se nessuno fa esperienza di quel calore; similmente, poiché il rosso di una superficie è identico al suo riflettere una certa lunghezza d'onda io potrei dire che una certa mela è rossa anche se nessuno la vede. Ora, supponiamo che il dolore sia identico ad un certo evento cerebrale come un'attività delle fibre C: potrei dire, quindi, che "c'è dolore" se c'è attività nelle fibre X anche se nessuno sente quel dolore? Ovviamente no: se non c'è esperienza del dolore non c'è dolore. Si noti che per Kripke, come per Ruth Barcan Marcus, l'identità è sempre necessaria, quindi il solo fatto che sia possibile attività nelle fibre C senza dolore (anche qualora fosse impossibile nel nostro universo) è sufficiente a negare l'identità.

Voglio però ragionare su un altro esempio al quale rifletto da tempo. È piuttosto banale, ma mi aiuta a capire quali tipi di identità tra mente e corpo possono essere sostenuti, e quali no. Immaginiamo che un certo stato cerebrale statico (non un processo), una certa distribuzione della materia nel mio corpo, sia identica ad una certa esperienza soggettiva. Vuol dire che ogni volta che il mio cervello è in quello stato, ogni volta che la sua materia è distribuita in quel modo, io ho una certa esperienza, diciamo per praticità che vedo rosso. Immaginiamo allora che io venga congelato di colpo, conservato nella grafite come Han Solo, e tutta la materia del mio corpo rimanga bloccata così com'è. Se quello stato fisico è identico al mio stato mentale, allora, finché io rimarrò bloccato in quello stato avrò costantemente l'esperienza della visione del rosso. Se dovessi rimanere congelato così per l'eternità, rimarrei bloccato per l'eternità nella mia visione del rosso. È corretto immaginare una cosa del genere?

Vi sono persone, morte, il cui cervello viene conservato; anche il cervello di Marvin Minsky, grande ricercatore nel campo dell'intelligenza artificiale, è stato estratto da suo cranio e crioconservato nell'attesa "transumanista" che si sviluppi la tecnologia necessaria a farlo risorgere, collegando il suo cervello a qualche computer magari. Supponiamo che il suo cervello, manipolato accidentalmente mentre veniva estratto, è rimasto congelato in uno stato associato ad una esperienza mentale. Dovremmo dire che lui ora, pur morto, sta aspettando qualcosa, e rimarrà così finché non verrà risorto o lasciato marcire? Supponiamo, per totale absurdità, che sia così. Ovviamente, essendo il suo cervello congelato, non sta creando nuove memorie, è fermo in uno stesso istante, e non si può neanche dire che abbia esperienza del tempo; ciò nonostante, ad ogni istante del tempo, finché il suo cervello rimarrà in quello stato, vi sarà anche un suo qualia sempre identico.

Io, quando dormo, credo di non avere alcuna esperienza, semplicemente non ho una mente. Immaginiamo, però, un altro caso: in realtà, io ho una mente anche mentre dormo, solo che la mia esperienza rimane fissa all'ultima esperienza fatta prima di addormentarmi, se io mi addormento vedendo il soffitto allora rimango tutta la notte bloccato nell'esperienza di quell'istante del soffitto, e non registro nessuna nuova esperienza fino al mattino.

[uno schema con cervelli sempre uguali, ma solo una esperienza all'inizio]

[faccio uno schema in cui, di istante in istante, un uguale cervello corrisponde ad una uguale esperienza]

Mi domando: potrei io distinguere tra i due casi?

Un ricercatore italiano, Giulio Tononi, ha creato una delle più ammirate teorie sulla coscienza umana. Contrariamente agli approcci "mainstream", Tononi ha scelto di partire dalla

fenomenologia, parte dall'intuizione che la mente esiste e ha certe caratteristiche normalmente riconosciute dai filosofi della fenomenologia (per esempio, che l'esperienza è un'unità, è specifica, che vi sono dei moduli, eccetera). La sua intuizione, allora, è che, se la mente sopravviene sul cervello, essa dovrebbe correlare con stati del cervello che hanno caratteristiche simili a quelle che riscontriamo nei fenomeni. L'idea, in sostanza, è che il livello di coscienza dipenda dalla "informazione integrata", la quale sale con l'aumentare dei rapporti di causa-effetto esistenti tra le varie strutture; cioè un ente è tanto più cosciente tanto è più grande la differenza tra l'informazione rappresentata dall'ente preso nel suo insieme e la somma dell'informazione rappresentata dalle sue parti prese in isolamento. In questo modo è possibile, in teoria almeno (si tratta di calcoli molto complessi), osservare un cervello e, in base alla sua struttura e attività, assegnare un certo grado di coscienza; così, poi, un computer può imparare a distinguere, osservando il sistema nervoso, una persona sveglia e distratta, una concentrata, una addormentata, un insetto, un sasso.

Ora, immaginiamo che, tramite questa teoria, si riesca a individuare un certo stato del cervello, un suo certo pattern di attività, che viene associato all'attività cosciente (almeno nel senso cognitivo, come stato di veglia, se non fenomenologico). Immaginiamo però che dei medici, esaminandomi, scoprano una caratteristica unica in me: il mio cervello, quando dormo, ogni venti minuti entra per un minuto in quello stato di coscienza; io però, di mio, non mi ricordo assolutamente nulla al mattino. Potrebbe trattarsi di un nuovo tipo di sonnambulismo, forse? Non è facile dirlo. Perché dovremmo credere che quello stato cerebrale sia di per sé cosciente? Perché viene normalmente osservato in persone sveglie, attente, attive, e ha le caratteristiche che associamo all'esperienza. Ma non dovremmo dire, invece, che proprio il fatto che si può osservare anche in una persona dormiente, e che questa non riporta alcuna esperienza neppure di sogno, sfa la teoria, dimostra che quello stato in realtà non è cosciente, o comunque non sufficiente alla coscienza? Certo, si può obiettare: era cosciente, è solo che il ricordo è stato cancellato prima del risveglio; ma perché questa interpretazione dovrebbe essere più sensata dell'altra?

È ovvio che io non posso distinguere tra i due casi sopra illustrati: io, forse, non ho avuto alcuna esperienza; forse, invece, ho avuto un'esperienza che, pur essendo soggettivamente durata un istante, è in realtà durata otto ore di orologio (è ovvio che il tempo soggettivo non è quello oggettivo, un'ora con la fidanzata non è percepita come un'ora seduto su una stufa bollente); in entrambi i casi al mattino penso "ho dormito" solo perché mi accorgo di colpo del cambiamento rispetto all'ultima esperienza che ricordo, che quest'ultima esperienza sia stata vissuta la sera prima per poi cadere nel vuoto, o sia stata vissuta solo un istante prima del risveglio, è irrilevante perché per me le due cose sono equivalenti. Stessa cosa per l'esempio di prima: forse sono stato cosciente e non ho ricordato, forse invece quello non è davvero uno stato cosciente, fare esperienze che non posso ricordare è come non farle.

Questo mi porta ad affermare che la memorizzazione è necessaria all'esperienza. Io, durante la notte, posso immaginare di fare sogni ogni microsecondo, solo che non li memorizzo, e questo, per me, è perfettamente equivalente al non sognare affatto. Ma la memorizzazione richiede cambiamento, è un processo che avviene nel tempo, richiede che, dalla zona del cervello che costituisce la mia esperienza presente, l'informazione venga in qualche modo registrata man mano che nuove esperienze presenti si susseguono. Questo mi suggerisce che nessuna "fotografia" statica del cervello, nessun modo di distribuire la materia nel mio cervello, può essere fatta corrispondere ad un'esperienza.

[nuovo schema in cui gruppi di istanti vengono fatti corrispondere ad una esperienza]

La coscienza così interpretata può ricordare concetti come quello di accelerazione. L'accelerazione indica un cambiamento nella velocità, e non c'è nessun singolo istante dell'universo che si possa dire manifestare l'accelerazione, l'accelerazione si può individuare solo confrontando diversi stati dell'universo, in diversi istanti; similmente, e molto banalmente, la presenza della mente non può essere determinata semplicemente osservando il cervello in un istante, bensì bisogna osservare come il cervello cambia nel tempo. Questo vuol dire che un identico stato mentale può essere associato ad una esperienza soggettiva, o anche no, a seconda di quali sono gli altri stati mentali che vengono prima o dopo: se io passo ad un certo stato mentale, e lo memorizzo, allora posso dire di avere esperienza di quello stato mentale, ma se il mio cervello è bloccato da otto ore non posso dire di avere davvero coscienza di qualcosa. Questo si collega, in modo interessante, anche a quanto detto parlando dell'identità: "io" sono "io" perché questo "io" non sta in una sostanza ma in un processo, in una relazione tra gli stati temporali che così formano un tutt'uno, "io" non sono presente in alcun istante ma emergo quando si confrontano gli istanti prendendoli come coppie di parti temporali; come ha senso parlare di accelerazione di una particella solo se a variare è la velocità di una stessa particella, così ha senso parlare di esperienza solo se sono sempre io ad avere certi stati cerebrali. Comunque, cosa ci dice questo sulla natura della mente? Stiamo partendo dal presupposto che la mente non sia una sostanza a parte, ma che sia sempre qualcosa di interno al mondo fisico perché così è più diretto spiegare la superveniencia: questo è confermato da quanto detto, perché la mente non si comporta come una sostanza autonoma, non è presente in nessun istante preso in isolamento. La mente, infatti, non superviene sulla sostanza fisica in sé, non superviene su sue proprietà intrinseche, ma bensì su proprietà estrinseche, relazionali, o meglio, superviene su suoi processi. Non si può quindi sostenere un'identità della mente con proprietà intrinseche della sostanza (né con la sostanza in sé): come l'accelerazione, la mente è un qualcosa che emerge considerando gli istante nel loro insieme. Questo è simile a ciò che si può affermare considerando la possibilità delle realizzazioni multiple (poiché diversi stati cerebrali corrispondono allo stesso evento mentale l'evento mentale non può essere identico a nessun stato cerebrale specifico), ma in quel caso si può affermare che l'evento mentale potrebbe essere identico ad una disgiunzione di stati ("o lo stato x, o lo stato y, o lo stato z..." andando avanti così fino a enumerare tutti gli stati funzionalmente equivalenti), mente in questo caso non si può: infatti nessuno degli stati presenti nella disgiunzione corrisponde di per sé ad alcuna sensazione, tutto dipende da come quegli stati (o anche la disgiunzione stessa) sono disposti nel tempo rispetto ad altri stati possibili.

La domanda che segue allora è: la mente può essere identica ai processi della sostanza fisica, cioè a proprietà fisiche relazionali? Questo è sicuramente vero se si parla della mente da un punto di vista cognitivo: il mio calcolare $2+2=4$ può essere identico ad un certo evento nel mio cervello, così come il calcolo della calcolatrice è identico ad un certo evento nei suoi circuiti. Questo però è insostenibile per la mente vista da un punto di vista fenomenologico: l'accelerazione è identica al cambiamento di velocità, ma i fenomeni non sono identici al cambiamento di struttura dei miei neuroni. Questo è implicito nella definizione delle proprietà fisiche: infatti le proprietà fisiche sono degli osservabili, ma l'esperienza in sé non è osservabile; inoltre pare necessario per fuggire alle argomentazioni anti-riduzioniste come quella di Kripke. Frege afferma, inoltre, che due termini non equivalenti possono riferirsi allo stesso ente solo se si riferiscono a sue diverse proprietà, come facciamo quando parliamo della statua e del suo marmo: al che, poiché è chiaro che l'esperienza del dolore non è equivalente alla descrizione di un certo processo neurale, dobbiamo dire che stiamo pescando proprietà diverse della sostanza fisica.

Non resta, allora, che sostenere che le proprietà mentali sono sì proprietà della sostanza fisica, ma non sono identiche ad altre proprietà fisiche.

Ciò che è fisico, è vero, non ha intrinsecamente proprietà mentali e proprietà fisiche, perché allora non sarebbe più fisicalismo; ciò nonostante, le proprietà fisiche solo tali che, in certe circostanze, relazionandosi in certi modi, creano anche proprietà mentali: due proprietà di una sola sostanza. Questo è quello che viene generalmente chiamato "dualismo delle proprietà", e, come il funzionalismo, non è in contraddizione con il fisicalismo, bensì lo integra e specifica. Dobbiamo passare allora alla causazione. È cosa comune parlare degli effetti di certe proprietà; è sospetto però interpretare una certa proprietà direttamente come una causa. La sostanza è, in generale, ciò che costituisce il livello più fondamentale della realtà: che siano particelle, idee, monadi o quant'altro è stato proposto nel corso della Storia. La sostanza è caratterizzata da certe proprietà, certi modi in cui è. Un elettrone ha certe proprietà, per esempio ha una carica, e ciò che ha carica può mollarci delle scosse; sarebbe strano però dire che la carica è la causa della scossa, o meglio, lo si può dire, ma ciò che si deve intendere è che l'elettrone è strutturato in modo tale da causare scosse: cioè, la proprietà non è una causa di per sé, ma dà alla sostanza certi poteri causali; se l'elettrone non avesse avuto la proprietà della carica non avrebbe potuto mollarci scosse, poiché la ha, ma è l'elettrone che ha questa facoltà, non la proprietà che non ha una esistenza autonoma.

Se si accetta questo modo di intendere la causazione, allora è ovvio che non può esserci sovradeterminazione in un fisicalismo emergentista. L'idea è che la sostanza sia caratterizzata da certe proprietà fisiche, le quali, in accordo con leggi fisiche (e sono leggi fisiche in quanto riguardano il comportamento della sostanza fisica e non altro), in determinate condizioni, possono far emergere nuove proprietà mentali le quali danno alla sostanza fisica nuovi poteri causali (in questo modo si può rendere il libro arbitrio intelligibile, se non altro, il che non vuol dire che il comportamento dei miei elettroni sia diverso da quello degli altri enti). Ciò che causa gli eventi futuri, però, è sempre la sostanza, non le proprietà: vi è causazione mentale, ma sempre e solo tramite la sostanza fisica. Vi è quindi chiusura causale fisica nel senso che ogni evento futuro è effettivamente causato in modo sufficiente dalla sostanza fisica, senza bisogno di alcuna altra sostanza; non vi sono invece "effetti emergenti" che supervengono su altri, effetti mentali totalmente separabili da quelli fisici, e quindi non c'è sovradeterminazione perché ogni evento futuro è causato dalla sola sostanza fisica, ha una sola causa e non due.

Un (quasi)fisicalismo

Tutte e teorie che abbiamo visto finora posso essere messe insieme per creare una complicata versione del fisicalismo:

- esiste - di rilevante per il nostro argomento - solo una sostanza, ed è di tipo fisico (questo è il fisicalismo)
- dalla sostanza fisica, che ha sempre proprietà fisiche intrinseche, possono emergere anche proprietà mentali (questo è il dualismo delle proprietà)
- questa emergenza può avvenire quando, in un certo ambiente, la sostanza fisica è strutturata in modo particolare, per esempio quando realizza - ad una adeguata scala di grandezza, rispettando determinati criteri per limitare i tipi di struttura accettabili - le funzioni di filtraggio e integrazione delle informazioni e autoanalisi (questa è una ipotesi di funzionalismo)
- le proprietà mentali non sono riducibili alle proprietà fisiche, la loro emergenza non può essere calcolata tramite le proprietà fisiche (emergenza forte, da cui il misterianismo)

Questa teoria sembra immune a pressoché tutte le obiezioni che abbiamo esposto finora, che sono anche le più famose e potenti della teoria della mente.

Non è soggetta alle argomentazioni anti-funzionaliste di Searle e Ned Block in quanto la funzione della struttura non è sufficiente all'individuazione di una mente: si può dire che la Cina disposta a cervello e la stanza di Searle cercano di far emergere la mente da proprietà che appartengono a persone e oggetti macroscopici, quando devono emergere da proprietà di particelle o comunque altre componenti; il caso del computer che si limita a rispondere a domande tramite un albero di risposte predefinite e l'assegnare funzioni diverse a seconda che a una martellata uno urla "ahi" o "d'oh" o che rida come un masochista, dipendono invece dal trovare i giusti limiti da imporre alla struttura (il primo ne pone pochi e deve richiedere che la struttura funzioni in un certo modo per esempio prevedendo elaborazione semantica e non solo sintattica, il secondo ne pone troppi e deve accettare che la funzione colleghi tipologie di eventi e non eventi singoli); infine il caso dei cervelli in vasca, cioè l'esternalismo semantico, è un problema se si crede che ogni stato cosciente sia necessariamente caratterizzato da intenzionalità cioè dall'avere un riferimento, ma c'è anche chi sostiene l'esistenza di qualia senza intenzionalità, che l'accoppiata qualia-riferimento è contingente, e quindi può ben essere che due stati mentali identici possono avere riferimenti diversi a seconda del loro contesto.

Non è soggetta neanche alle argomentazioni anti-fisicaliste in quanto accetta che i qualia siano proprietà distinte dalle fisiche e non riducibili a queste: è ovvio allora perché non capiamo cosa voglia dire essere un pipistrello, perché Mary la scienziata dei colori impara qualcosa di nuovo quando vede i colori (le proprietà mentali non sono identiche alle fisiche e la loro emersione non è prevedibile a partire dalle parti), e perché possiamo accettare gli zombie filosofici come logicamente possibili senza problemi (gli zombie filosofici mostrano che potrei creare un duplicato fisico del nostro universo senza per questo dover immaginare che gli uomini di quell'universo abbiano una mente, ma questo non vuol dire che la mente sia una sostanza non-fisica, vuol dire solo che sto copiando le proprietà fisiche della sostanza fisica ignorando le sue proprietà mentali). L'unico caso che la teoria non può accettare è quella dei qualia invertiti, se il mentale emerge dal fisico allora da uno stesso stato fisico non possono emergere più stati mentali (vorrebbe dire che quale stato mentale emerge dipende dal caso, e chiaramente non è così); si può dire, allora, che un caso del genere è una pura e semplice impossibilità metafisica in tale teoria. Questo va riconosciuto anche per gli zombie filosofici: pur logicamente possibili, la sostanza fisica è tale che non può non far emergere anche proprietà mentali emergenti; inoltre, anche se non è necessario che i qualia siano di per sé causalmente efficaci in modo diretto, se accettiamo che la mente in generale è causalmente efficace allora è chiaro che uno zombie filosofico senza mente non potrà mai comportarsi proprio come mi comporto io.

Una critica rilevante e nuova è che gli zombie filosofici sono stati esclusi dal problema in modo troppo semplice, così semplice che, in effetti, la teoria non sembra più una forma di fisicalismo. Per il fisicalismo, infatti, tutto ciò che esiste è solo il fisico, ma questa teoria prevede proprietà mentali totalmente distinte dalle proprietà fisiche.

In effetti, la definizione di fisicalismo dalla quale siamo partiti è molto liberale: tutto ciò che è richiesto è che il mentale supervenga sul fisico, e così è anche con proprietà emergenti in modo forte (infatti non c'è cambiamento mentale senza che ci sia un cambiamento nella sostanza fisica). Molti filosofi però rifiutano questa definizione ritenendola insufficiente al fisicalismo: ciò che richiedono è non solo che esista solo una sostanza fisica, ma anche che questa abbia solo proprietà fisiche. A questo punto la domanda è: cos'è una proprietà fisica? E questo è un bel problema, perché, mentre la definizione di "fisicalismo" è stata discussa in tutti i modi, la definizione di "proprietà fisica" rimane piuttosto vaga e dubbia, e questo nonostante il ruolo delle proprietà fisiche e non-fisiche nel fisicalismo sia importantissimo e molto discusso.

D'altro canto, una buona fetta dei filosofi pare riconoscere come proprietà fisiche non solo quelle discusse dalla scienza fisica, ma anche quelle che supervengono su queste; e se è così non c'è motivo per il quale non potremmo dire che le nostre proprietà mentali sono in realtà fisiche in virtù del solo fatto che il mentale superviene sul fisico. Se per il fisicalismo un mondo identico al nostro in tutti i fatti fisici lo è in tutti i fatti, be', così è anche per questa teoria: poiché, poste tutte le proprietà fisiche della sostanza fisica e tutte le leggi fisiche, le proprietà mentali emergono automaticamente, un mondo fisicamente identico a questo è anche mentalmente identico a questo, se un ente è fisicamente uguale è anche impossibile che non abbia una mente. Perché dunque non dovrebbe essere un fisicalismo, questo? Se non va bene questa definizione di fisicalismo, basata sull'emergenza, l'alternativa pare essere solo la teoria dell'identità, la quale è oggi più difficilmente sostenibile che in passato.

Alla definizione proposta si può sollevare un'obiezione simile a quelle già viste per il funzionalismo: per esempio, si potrebbe immaginare un universo uguale al nostro in tutto, con l'unica differenza che esiste un atomo d'idrogeno in più negli anelli di Saturno; questa differenza, pur insignificante, rende quel mondo un non perfetto duplicato fisico del nostro, e legittima l'ipotesi di fatti mentali totalmente diversi dai nostri, anche se ciò è controintuitivo. Volendo, però, si può accettare tale osservazione: un universo che prevede una molecola in più, per quanto possa sembrare una differenza minimale, può richiedere delle condizioni di partenza molto diverse, per esempio delle leggi diverse dalle nostre, e quindi può prevedere anche diversi fatti mentali; oppure, invece di parlare di un duplicato fisico del mondo, si potrebbe parlare semplicemente di un duplicato fisico del mio corpo, o del mio cervello, dove la presenza di una molecola in più potrebbe anche essere rilevante.

Penso però sia corretto individuare nel dualismo delle proprietà con emergenza forte il punto più debole della tesi. Le proprietà fisiche e le proprietà mentali sono enormemente diverse: la proprietà fisica degli eventi elettrochimici non ha proprio nulla della qualità di una sensazione soggettiva. È chiaro che questa differenza è una differenza di prospettiva, ciò che in terza persona appare come un evento elettrochimico in prima persona appare tutt'altro; ma com'è possibile che sia così, che una stessa cosa possa manifestare proprietà forse addirittura contraddittorie, non è affatto chiaro. L'idea dell'emergenza forte sembra proprio una fuga, una rinuncia, e una petizione di principio: si pone, come posit, che il mentale possa emergere dal fisico, ma come ciò sia possibile ci si rifiuta direttamente di spiegarlo. Se è legittimo fare così, allora è altrettanto legittimo agire in senso contrario, e porre come posit che non è possibile che il mentale emerga dal fisico. Possiamo allora considerare altre possibilità...

Dualismo

Abbiamo posto, per ipotesi, che il mentale non possa emergere dal fisico. D'altro canto, l'esistenza del mentale è ovvia, è l'unica cosa della quale possiamo essere certi, è l'esistenza del fisico che è dubbia, quindi non possiamo neanche dire che esistano solo proprietà fisiche. Una soluzione, allora, può essere il passare da un dualismo delle proprietà direttamente ad un dualismo della sostanza: materia e mente sono due sostanze totalmente diverse e separate, ognuna con le sue proprietà.

Questa idea è notoriamente attribuita a Cartesio, tanto che normalmente si parla di "dualismo cartesiano", e si pensa generalmente che nasconda motivazioni religiose, che sia sostanzialmente una scusa per giustificare la possibilità della sopravvivenza dell'anima separatamente dal corpo. In questa pagine, però, ho mostrato spesso che le idee che sembrano più vicine al rigore scientifico sono invece spesso sostenute da filosofi relativisti, mentre quelle che sembrano più vicine all'ideologia religiosa sono invece molto distanti da ciò che la teologia crede normalmente. Anche questo è uno di quei casi.

Nella filosofia medievale il concetto di "mente" appare molto diverso da quello che noi abbiamo oggi. In primo luogo, San Tommaso parla di anima e corpo. È vero che sono due sostanze diverse, ma credere che siano separabili senza problemi è un errore: anima e corpo formano un sinolo, la persona è unità di anima e corpo, un corpo senz'anima non è una persona, ma un'anima senza corpo neppure lo è. Alcuni tra i primissimi padri della Chiesa (parliamo quindi di pensatori del primo e secondo secolo) erano addirittura convinti che, in effetti, l'anima non potesse sopravvivere senza corpo, e che se di fatto sopravviveva era solo per una grazia di Dio, per un miracolo; un'idea, questa, che oggi sta tornando in voga tra i teologi ortodossi, tra l'altro. Non vi è dunque un dualismo tale da giustificare la separabilità dell'anima dal corpo, neanche nel senso che abbia una naturale facoltà di sopravvivenza. Ma attenzione, perché non ci sono solo anima e corpo: c'è anche lo spirito; spirito vuol dire respiro, e le sue proprietà sono effettivamente simili a quelle di un gas, nel senso che non è un ente immateriale come l'anima, ma non è neanche un ente materiale come il corpo, è un qualcosa che sta a metà via, ha una massa ma pochissima.

Tra anima, corpo, e spirito, dove sta la mente? In effetti, in nessuno dei tre. È vero, Tommaso parla dell'anima in un modo che le assegna delle facoltà che noi attribuiamo alla mente, come il ragionamento. D'altro canto, Tommaso attribuisce facoltà mentali anche al corpo: per esempio, per lui la memoria, la memoria derivata dai sensi almeno, è una cosa che riguarda il corpo, e che l'anima non ha. Quindi la mente come noi la intendiamo non sta né solo nel corpo, né solo nell'anima, e sarebbe un errore imbarazzante il finire per credere che l'anima sia la mente, e che la mente così come noi la intende la filosofia moderna possa sopravvivere alla morte del corpo con tutte le sue facoltà (il che non vuol certo dire che i teologi medievali credessero nel sonno dell'anima come i testimoni di Geova). E questo errore, in effetti, lo fa Cartesio; ma non è proprio un errore: è proprio un cambiamento di categorie, un passaggio da ciò che credeva la teologia medievale ad un'idea nuova, quella che la mente sia l'anima e che sia del tutto indipendente dal corpo.

E lo spirito? Questo in Cartesio ritorna, e gli è necessario. Vi era infatti un ovvio problema di causazione: è ovvio per noi come per lui che la mente ha un potere causale, determina dei comportamenti; ma se la mente è astratta, e il corpo è materiale, come può la mente avere un potere causale sul corpo? Ebbene, è vero che lo spirito ha poca energia, ma un po' lo ha: la mente lo controlla, e quando la materia è abbastanza rarefatta (come un gas appunto) lo spirito è abbastanza energico da riuscire a influenzarla. In un certo senso, si potrebbe dire che Cartesio credeva che l'anima influenzasse il comportamento facendo entrare aria nel cervello...

Questa teoria ci mostra subito il più grande punto debole del dualismo: la "chiusura causale" del fisico. La termodinamica è chiara: l'energia si conserva, può trasformarsi, l'energia può diventare lavoro e il lavoro generare calore, ma non si crea e non si distrugge; c'è un valore numerico, insomma, che, per quante cose si possano fare nell'universo, non cambia mai, non aumenta mai, non cala mai. Ora, se la mente è astratta sta al di fuori dello spaziotempo, al di fuori dell'universo materiale; se tale mente può interagire col mondo fisico, allora deve per forza farlo avendo degli effetti energetici: se un atomo si sta muovendo ad una certa velocità in linea retta la mente non può farlo curvare, o fermarlo, se non immettendo energia per "spingerlo". Ma se una cosa del genere fosse anche solo possibile, be', sarebbe evidente: la quantità d'energia nell'universo aumenterebbe. Forse non è necessario, forse la mente può in qualche modo "spostare" energia da un sistema all'altro: ma, di nuovo, anche ammesso sia possibile un qualche tipo di causazione senza variazione nell'energia, se fosse così sarebbe evidente, lo vedremmo, vedremmo un sacco di eventi inspiegabili, particelle che scompaiono di colpo per fermare altrettanto di colpo un gruppo d'altre, ma noi non osserviamo nulla del genere. Insomma, tutto funziona proprio come se le relazioni causali si dessero solo all'interno

dell'universo fisico, solo tra sostanze fisiche, e non pare proprio che la mente possa avere alcun effetto su di esse.

Si potrebbe provare a sostenere, allora, che il dualismo pare essere richiesto da quanto detto sul libero arbitrio: il mio cervello fisico, in quanto fisico, pare totalmente determinato, e se posso davvero pensare e ragionare in un modo almeno un poco libero devo avere delle componenti che vadano oltre il fisico. Non è però necessario che sia così, il nostro universo fisico potrebbe benissimo funzionare in un modo tale da non negare la possibilità del libero arbitrio, certo non capiamo come ciò sia possibile, magari lo capiremo in futuro, magari nonostante la nostra libertà non potremo scoprirlo, ma comunque nulla ci obbliga a considerare la natura fisica sempre e solo come determinata; al massimo, si può sostenere che il dualismo rende molto più facile spiegare com'è che la libertà è possibile, ma nulla di più.

D'altro canto, se è vero che quest'argomentazione non ci dà motivo di credere che sia necessario postulare ulteriori sostanze per rendere la nostra visione coerente, è anche vero che non abbiamo alcun motivo di credere che esista solo un'unica sostanza e che sia fisica. Ci è chiaro che noi possiamo acquisire reale conoscenza, ma ci è anche chiaro che siamo esseri limitati, che come ci sono cose che sono al di là della comprensione di un cane così ci sono cose che sono al di là della nostra comprensione, i teoremi di Godel e l'indeterminatezza della fisica quantistica sono esempi celebri perché ci fanno rendere conto che, un poco paradossalmente, già sappiamo che ci sono cose che non sapremo mai. Ora, il nostro universo ci pare, grosso modo, piuttosto comprensibile: non lo capiremo mai nella sua interezza, ma la scienza ci ha portato a comunque a comporre un suo ritratto che pare decente. Ma: da un lato, sarebbe strano se tutto ciò che esiste fosse qualcosa di per noi comprensibile e facilmente indagabile, è ben probabile che vi sia molto che ci sfugge; dall'altro, se già adesso faticiamo a capire il comportamento del nostro universo, be', potrebbe benissimo esserci molto altro, al di là di questo universo, che ci è semplicemente inimmaginabile. Proprio Godel, in uno strano riassunto della sua filosofia, esprime una serie di visioni di questo tipo: "esistono altri mondi ed esseri razionali, della diversa e più alta tipologia", "il mondo in cui viviamo ora non è l'unico in cui viviamo o abbiamo vissuto". Certo, sarebbe sbagliato dire che possiamo in qualche modo descrivere universi paralleli, paradisi, inferni, paesaggi di mondi possibili; non è però assurdo dire che l'esistenza di altre sostanze, oltre quella fisica del nostro universo, può essere considerata probabile.

Monismo neutrale

Questo apre ad una nuova possibilità: se è vero che le proprietà mentali non possono emergere dalle fisiche, ma è evidente che esistono entrambi i tipi di proprietà, né possiamo ricondurle a due sostanze distinte, allora non possiamo che negare direttamente l'idea che la sostanza che compone il nostro mondo sia semplicemente una sostanza fisica: è, bensì, un altro tipo di sostanza, una sostanza psicofisica, che ha un tipo misto di proprietà, proprietà psicofisiche, le quali possono essere descritte sia da una prospettiva strettamente fisica, sia da una strettamente mentale, anche se di per sé non sono nessuna delle due cose.

Questo non deve essere inteso come il fatto che ogni ente dotato di proprietà materiali, foss'anche un elettrone, ha anche proprietà mentali nel senso che ha esperienze. Questa ipotesi, detta "panpsichismo" ha invero avuto anche i suoi sostenitori; anche la teoria di Tononi esposta sopra, rappresentando la coscienza come un continuum, implica una sorta di panpsichismo (si potrebbe dire che tutto è più o meno cosciente, anche i sassi sono coscienti, semplicemente lo sono molto meno di noi). Però quasi nessuno, oggi, affermerebbe qualcosa del genere in modo letterale. Ciò che dobbiamo dire, casomai, è che certi enti, come i sassi, oltre alle proprietà fisiche, hanno proprietà "protopsichiche", non hanno esperienze ma hanno un altro tipo di

proprietà che può diventare esperienza cosciente, hanno per così dire le componenti della coscienza.

La posizione del monismo neutrale ha avuto diversi sostenitori illustri nel corso della storia, e pare una posizione di buon senso. Io sono stato uno zigote unicellulare, una gastrula, un esserino di poche cellule, di centinaia, migliaia di cellule, finché non sono nato: quand'è che la mente è apparsa? Se non c'è emergenza di nuovi tipi di proprietà, allora le proprietà che ho io devono essere dello stesso tipo di quelle che avevo quando ero anche una singola cellula, quindi, poiché è ovvio che io ho proprietà mentali, devo rispondere che le ho sempre avute, e anche i miei atomi le hanno, devo dire addirittura che le proprietà mentali (o proto-mentali) sono alla base della realtà.

Che questa sia una corretta rappresentazione del monismo neutrale è dubbio tanto quanto lo è la rappresentazione che ho dato del fisicalismo. Sto definendo il monismo neutrale come la posizione che prevede proprietà mentali e fisiche come di pari livello, ugualmente basiche (cioè vi è una realtà di fatti basici sia mentali che fisici) o ugualmente emergenti (sia il mentale che il fisico è composto da enti neutrali); non considero monismo neutrale invece la posizione che prevede solo una delle due proprietà come emergente. Non tutti sono d'accordo su questo: per alcuni, affinché una certa posizione sia un monismo neutrale, è sufficiente che uno stesso ente costituisca sia fatti fisici che mentali; così, il quasi-fisicalismo proposto sopra sarebbe in realtà un monismo, ma anche un fisicalismo riduzionista in cui tutti i fatti mentali emergono debolmente dai fisici può essere considerato un monismo: se i fatti fisici sono tali da comporre i mentali allora tali fatti fisici sarebbero considerati contemporaneamente fatti neutrali.

Volendo, si può dire così allora: che vi sono posizioni che possono essere considerate contemporaneamente fisicalismi e monismi neutrali.

Per fare un esempio, prendiamo il monismo sostenuto da Russell (uno dei tre leggendari fondatori della Sacra Scuola del Neutralismo, assieme a William James e al fisico e psicologo Mach). L'idea di Russell è che la fisica descriva solo la struttura causale della materia, cioè che descriva solo le proprietà intrinseche, relazionali, della materia: un elettrone, per esempio, è definito da una serie di formule che descrivono il suo comportamento, l'elettrone è definito dalla sua capacità di attirare i protoni e respingere altri elettroni. Se è così, allora la fisica non solo è una descrizione incompleta del mondo, ma è addirittura una descrizione incompleta anche della sola materia: non ci dice niente sulle sue proprietà intrinseche che pur deve avere se esiste. Al che, invece di ipotizzare che la mente sia una sostanza ulteriore, la soluzione economica di Russell è immaginare che le proprietà mentali siano le proprietà intrinseche degli enti fisici: si potrebbe dire che un elettrone ha le proprietà estrinseche descritte dalla fisica ma in più ha anche delle proprietà intrinseche dalle quali derivano le proprietà che invece conosciamo (sia pur molto alla lontana) tramite la psicologia. Orbene, mentre il fisicalismo emergentista viene a volte presentato come un monismo, quello di Russell è al contrario, un monismo che nella letteratura accademica viene spesso presentato anche come un fisicalismo. Il problema principale della teoria così com'è stata definita è che risulta incredibilmente oscura. Cosa sarebbe, esattamente, questa sostanza "psicofisica"? Cos'è una proprietà "psicofisica" esattamente? Capisco come le proprietà fisiche interagiscono tra loro, capisco come due atomi si legano a formare una molecola; ma dal punto di vista mentale, cosa dovremmo dire? Com'è che due "atomi mentali" formano una "molecola mentale"? Quante sono le proprietà mentali? Ogni evento mentale può essere fatto corrispondere ad una proprietà a sé, oppure no? Mentale e fisico paiono due tipi di proprietà estremamente diversi, non hanno pressoché nulla in comune: com'è possibile, quindi, che qualcosa sia "psicofisico", com'è possibile che tenga insieme proprietà così contraddittorie? Se il comportamento della parte fisica è spiegata da leggi fisiche, dovremmo aspettarci che questa sostanza risponda per la sua parte mentale anche

a leggi psicologiche? Dovremmo credere all'esistenza di due tipi di leggi, che lavorano contemporaneamente? Com'è possibile che esperienze così diverse (la visione di un paesaggio non ha nulla dell'odore del cioccolato) siano tutte riconducibile ad uno stesso principio psicofisico? E se non è un stesso principio, be', quanti ce ne sono? Uno per ogni sensazione? Sembra, insomma, una teoria che lascia molte più domande di quelle a cui cerca di rispondere. Una possibile difesa, pur estrema, è rinunciare all'idea d'ispirazione russelliana che le proprietà mentali siano intrinseche: tutto è solo una proprietà estrinseca. In questo modo si può dire che molte di queste domande non necessitano di risposta: ciò che è essenziale da conoscere di un certo ente è il suo comportamento, se io so che un elettrone attira i protoni e che combinato in certi modi crea i fenomeni allora non c'è null'altro che ho bisogno di sapere. È una risposta corretta, ma che lascia l'amaro in bocca: un fisico di inizio Novecento può accontentarsi di dire che l'elettrone è ciò che attira il protone senza chiedersi esattamente come faccia a fare ciò (si dirà "è una sua proprietà, funziona così il mondo"), ma per un filosofo invece questa è una domanda legittima.

Qualia e linguaggio

Nel monismo l'analisi linguistica-concettuale è particolarmente importante: noi distinguiamo mentale e fisico tramite il linguaggio, la distinzione è il frutto dell'applicazione di nostri schemi concettuali ad una realtà che di per sé sarebbe indistinta.

Un esempio di teoria ad alto contenuto linguistico è quella del "monismo neutrale" di Donald Davidson. Davidson parte da tre principi:

- che esista un rapporto causale tra mentale e fisico
- che ogni rapporto di causazione è mediato da leggi
- ma che non esistono leggi psicofisiche in grado di legare eventi mentali e fisici

Si può partire dallo scetticismo di Hume per i rapporti di causa, l'idea che non ci sia alcun fatto che costituisca la causazione. La soluzione scettica di Hume è che possiamo parlare di causazione non prendendo eventi singoli, ma tipologie di eventi nel loro complesso. Davidson intende le leggi in questo modo, come ricorrenze che legano tipologie di eventi in modo tale che, se si dà uno, si dà sempre anche l'altro. Una legge psicofisica, quindi, non dovrebbe legare semplicemente uno specifico evento mentale ad uno specifico evento fisico, ma tipi di eventi fisici a tipi di eventi mentali in modo esatto e preciso. Tramite la teoria della traduzione di Quine e soprattutto tramite un complessissimo ragionamento sul concetto di razionalità Davidson cerca di mostrare che ciò è impossibile; cerca di dimostrare, cioè, che è impossibile trovare vere leggi psicologiche, che è impossibile descrivere la mente (o il comportamento in termini mentali, come può fare la psicologia, la sociologia, la storia, l'economia) tramite leggi che siano prive di eccezioni. Se ciò è vero, se non ci sono leggi psicologiche, allora non può esserci neanche rapporto di causazione tra mentale e fisico; ma se il mentale è identico a qualcosa di fisico, allora ecco che i tre principi possono ora stare insieme.

Davidson non fa mai ontologia, è neutrale, per lui è "fisico" ciò che è descritto in termini fisici e "mentale" ciò che è descritto in termini mentali: vi è quindi un dualismo che però è puramente linguistico, riguarda il linguaggio e non il mondo. Nota così che si può dire che è fisico proprio ciò che è legato da rapporti di causa; quando però l'evento viene descritto in termini mentali il rapporto di causa non è più individuabile. Si può affermare, insomma, che esiste una sola sostanza, ma che quando questa viene descritta in modo mentale non risulta riducibile al fisico; quando parliamo di causazione tra mentale e fisico, per esempio quando

diciamo che il dolore causa l'urlo, possiamo farlo in realtà solo intendendo una causazione da sostanza fisica a sostanza fisica, e non da mentale a fisica.

La teoria di Davidson è particolare e ha vari pregi: tenta infatti di derivare la sua teoria in modo totalmente a priori, senza alcuna necessità di prove sperimentali; così può affermare che il fatto che non ci sia che il fisico, che il mentale sia solo una interpretazione, è una necessità metafisica, non solo una teoria. Questa necessità metafisica è data proprio dalla causalità: gli eventi mentali hanno potere causale, ma solo gli eventi fisici hanno potere causale, l'evento mentale è ciò che realizza ciò a cui si riferisce, e poiché gli eventi mentali si riferiscono a qualcosa con potere causale devono riferirsi a qualcosa di fisico. Per di più, il problema mente corpo, in questo stile, si può eliminare facilmente: si può dire che il linguaggio mentale non si può usare per descrivere fatti neurologici, e che il linguaggio neurologico non si può usare per descrivere esperienze mentali, e quindi che la divisione tra mente e corpo è creata al modo in cui si crea la divisione tra la statua e il suo marmo.

Questo non vuol dire, però, che a così grandi pregi non si oppongano altrettanto grandi difetti. Noi accettiamo come evidente che la mente supervenga sul corpo; ciò vuol dire che ogni volta che si dà un certo cambiamento mentale si dà anche un certo cambiamento corporale; se si usa la definizione di legge usata da Davidson, allora risulta che, se c'è supervenienza, c'è anche un collegamento causale tra mente e corpo: a che, o la teoria rinuncia al terzo principio (e viene distrutta), o nega la supervenienza (che sarebbe assurdo). La teoria, inoltre, è sempre sull'orlo dell'epifenomenalismo: se non ci sono leggi che legano mentale e fisico, in che modo la mente dovrebbe avere potere causale? Appunto, non è la mente che causa, è la sostanza fisica che viene interpretata in termini mentali, ma perché allora non dovremmo dire direttamente che la mente non ha alcun potere causale? Infine, la teoria della causazione che usa non è affatto l'unica possibile. Così, nessuno dei tre principi risulta ben giustificato.

Qualia e linguaggio privato

La più importante argomentazione linguistica in merito alla mente, però, si trova in Wittgenstein.

Wittgenstein non apprezzava molto tutto il discorso sul problema mente-corpo: gli sembrava (come con ogni problema) che fosse tutto un fraintendimento linguistico. Non credendo alla possibilità di un linguaggio privato, non credeva neanche alla possibilità dei qualia intesi come esperienze comprensibili solo a chi le vive. Usa l'esempio di una comunità nella quale ognuno ha una scatola con dentro uno "scarafaggio", ma ognuno può guardare solo dentro la propria scatola: in questa comunità la parola "scarafaggio" avrà un uso molto dubbio perché nessuno può sapere a cosa si riferiscano esattamente gli altri; similmente, con le esperienze mentali, contrariamente a quanto accade col significato delle altre parole, nessuno può correggermi, io non posso dire che darò un certo significato al termine finché gli altri parlanti non mi daranno motivo di dubitare della correttezza di ciò perché nessuno può farmi dubitare, se il significato della parola è l'uso io non posso in questo caso controllare l'uso che gli altri fanno della parola né come reagiscono al mio uso. In sostanza, i qualia risultano indiscutibili per Wittgenstein. Vi sono in effetti "estremisti", come i Churchland, che sostengono una versione "eliminativista" o "revisionista" del fisicalismo: l'idea è che certi termini, quelli della psicologia "popolare" nello specifico, sono semplicemente inadatti a parlare della mente, e per questo devono essere eliminati man mano che la neurologia si evolve; così come in passato parlavamo di certe malattie mentali tramite le categorie della demonologia, per esempio invocando la possessione, ora parliamo di stati mentali soggettivi come desideri e intenzioni, ma come abbiamo eliminato il linguaggio della demonologia, in futuro elimineremo il linguaggio mentale, e potremo finalmente esprimerci totalmente in termini oggettivi. Di nuovo,

sembra una posizione scienziata, e invece le sue radici stanno nei pensatori più postmoderni possibili, come Feyerabend e Rorty, che descrivono la scienza come se fosse poco più di un gioco e un'opinione tra le tante possibili.

Putnam affronta questo problema a più livelli.

Anzitutto, analizzando il concetto di razionalità e i punti deboli del principio di verifica dei positivisti logici (quella norma anti-realistica per la quale è sensato solo ciò che può essere dimostrato come vero) giunge alla conclusione che "non ci si può servire di regole pubbliche per decidere che cosa si possa discutere e giustificare razionalmente in filosofia". È ovvio che affermare "è razionale solo ciò che è verificato tramite il metodo xy" è autorifiutante se l'affermazione stessa non è verificata tramite il suo metodo; questo, però, vuol dire anche che dal fatto che la mia cultura presenti il metodo scientifico come esempio di razionalità non posso affermare che è razionale ciò che segue il metodo scientifico (a meno che io non voglia affermare qualcosa di irrazionale), non posso dedurre la razionalità semplicemente da ciò che è presentato come tale. E questo è un gran punto: vuol dire che dall'uso della parola "razionale" non posso derivare una esposizione del suo significato.

È ovvio che se riproponiamo, per la razionalità, il ragionamento fatto dagli eliminativisti per i qualia, e diciamo "la correttezza del nostro uso del concetto di razionalità non può essere controllato tramite l'uso pubblico, quindi è insensato, non c'è alcuna razionalità", be', affermiamo qualcosa di paradossale.

Passando al concetto di incommensurabilità sostenuto da Kuhn e Feyerabend affronta il problema del linguaggio privato in modo diretto: la loro idea infatti è proprio che il concetto di razionalità e giustificazione sia totalmente e irriducibilmente privato in quanto il significato dei termini utilizzati nelle teorie scientifiche cambiano di epoca in epoca, non possiamo dire che i dati che osserviamo oggi confermano la teoria della relatività di Einstein perché per Einstein la teoria aveva un significato diverso da quello che ha per noi, stiamo confermando un'altra teoria. Ora, vi è in filosofia il "principio di carità", per il quale dobbiamo sforzarci di interpretare le affermazioni altrui nel modo più vantaggioso, non dobbiamo scartarle alla prima obiezione ma bensì dobbiamo comportarci come se il nostro avversario intellettuale intendesse la sua teoria in un modo tale da poter superare quell'obiezione (se esiste tale modo); se non facessimo così, infatti, potremmo anche affermare direttamente che il nostro interlocutore dice cose "neppure false" in quanto sta semplicemente emettendo rumori a cui non vuole assegnare alcun significato. Questo, però, è proprio quello che dovremmo dire se la tesi dell'incommensurabilità fosse vera senza alcun limite: poiché ognuno intende i termini come vuole, in modo privato, per noi è proprio come se non dicessero niente, noi possiamo comprendere il senso solo dei nostri propri termini. Questo è assurdo: noi riteniamo che gli altri esseri umani siano, be', esseri umani, proprio come noi, e che abbiano una mente come noi, che siano razionali come noi, che siano coscienti come noi; questo però necessita che gli si passino anche i nostri concetti, in modo tale che sia possibile in linea di principio capirci. Putnam riconosce che, essendo i qualia privati, non possiamo sapere se una pietra, la Cina, la stanza di Searle, hanno una mente; ciò rende legittime sia una interpretazione del mondo in cui la hanno, sia una in cui non la hanno. La tradizione, però, ci ha trasmesso schemi per i quali certi enti sono coscienti, e altri no; in futuro, dovessimo lavorare fianco a fianco con robot capaci di interazione intelligente, le nostre sensazioni potrebbero cambiare e la nostra cultura potrebbe evolversi in modo tale da permetterci di assegnare una mente anche a dei robot, e forse anche a sassi, montagne, nazioni, squadre di calcio e quant'altro. In questo senso, il problema della mente non riguarda qualcosa di nascosto, riguarda il comportamento manifesto e le intuizioni che causa in noi.

In sostanza, arriviamo ad individuare un particolare criterio di somiglianza: noi non possiamo sapere se siamo oppure no gli unici esseri viventi ad avere esperienze coscienti, l'amico mi dice che lui ha esperienze mentali ma non so se sia vero o se lo stia semplicemente dicendo (in fondo, anche non le avesse, mi direbbe comunque ciò che il suo cervello lo spinge a dire); di fatto, però, noi ci basiamo sulla somiglianza apparente, poiché gli altri umani sono molto simili a me in tutto ciò che è esterno e osservabile, allora ipotizzo (legittimamente) che siano come me anche per i fatti interiori, e quindi che abbiano esperienze coscienti che siano molto simili alle mie, e concetti delle parole simili ai miei. Non bisogna fare l'errore di pensare che questo voglia dire che "poiché gli altri umani sono simili a me in quanto a corpo, lo sono anche in quanto a mente": non si parla infatti di corpo, quanto di comportamento; io assegno una mente agli altri uomini e animali perché, partendo dal comportamento, cerco di abduire le cause mentali di questo, e creo così in me una simulazione della mente altrui; se riesco a fare ciò in modo efficace (e ci riesco se il comportamento che osservo è simile al mio, è qualcosa che posso comprendere) allora assegno una mente a quell'ente.

Ora, questo vuol dire che quando noi parliamo dei qualia, ci stiamo effettivamente riferendo a qualcosa di interno e privato. Questo, però, non implica automaticamente che sia impossibile un controllo dell'uso del termine in quanto partiamo dal presupposto che gli uomini siano simili tra loro, che tutti abbiano effettivamente dentro la scatola uno scarafaggio proprio come il nostro, e quindi che il riferimento del termine sia effettivamente lo stesso e sia discutibile: si può affermare che questo presupposto non è più legittimo di qualsiasi altro, ma se non supponessimo così staremmo direttamente negando l'umanità degli altri uomini. Come detto nella prima sezione, i nostri termini potrebbero effettivamente riferirsi solo a esperienze soggettive, il fatto però che queste esperienze soggettive abbiano un collegamento forte con il mondo oggettivo (com'è, per esempio, se le nostre esperienze supervengono su caratteristiche del corpo) ci mette comunque nella posizione di poter discutere in modo significativo, si aggiunge semplicemente un passo in più in quanto, invece di riferirci direttamente a fatti oggettivi, ci riferiamo a fatti oggettivi tramite una esperienza soggettiva.

Quest'ultima affermazione trova un controesempio problematico in Donald Davidson.

Davidson immagina di essere sdraiato sotto un albero morto, in una palude. Un fulmine lo colpisce, viene smolecolarizzato, ma, per pura coincidenza, il fulmine porta altre molecole presenti nell'albero e nell'aria a ricombinarsi in modo tale da comporre un nuovo essere, Swampman, proprio uguale a lui, che fa tutto ciò che farebbe lui, dice tutto quello che farebbe lui, ma non è lui. L'intuizione di Davidson è che questo nuovo essere non avrebbe una mente nel senso che non avrebbe intenzionalità, non avrebbe la capacità di riferirsi ad enti esterni (come la stanza di Searle o il computer di Block) in quanto manca quella catena causale che costituisce il collegamento tra il proprio pensiero e l'ente a cui si riferisce (per l'esperimento mentale della Terra gemella, infatti, il referente della parola "acqua" è la sostanza che "causa" l'uso di quella parola). Ora, l'esperimento, ben più che dirci qualcosa sulla mente, sembra dirci qualcosa sulla semantica: l'intuizione più forte è che Swampman abbia una mente (anche se non la mia, in effetti), che abbia intenzionalità, e che quindi l'esternalismo semantico sia sbagliato.

Senza dubbio il prevedere una connessione causale tra le mie parole e i loro referenti spiegano come io riesca a riferirmi a qualcosa in circostanze normali, quando ho la possibilità di avere un certo oggetto sotto gli occhi. Non è però sempre necessaria la presenza di tale connessione. Immaginiamo che io prendo un libro di un certo autore, Mario Rossi, e che sul retro ci sia una sua foto. Io posso indicare la foto e dire "quest'uomo": io mi riferisco alla mia esperienza, causata dalla foto, la quale mostra un certo uomo a causa del suo essere davanti all'obiettivo, e in questo modo la catena causale mi porta al referente che io intendo. Supponiamo però che,

come nel caso di Luther Blisset, ad un certo punto io venga a scoprire che l'autore del libro non esiste, il libro è stato scritto da una pluralità di individui, e anche la foto è un collage di elementi di foto diverse. A questo punto la catena causale mi porta, dalla foto, al lavoro di un qualche grafico che ha mescolato i tratti per creare questo nuovo volto: dovrei dunque dire che è questo il mio referente? Dovrei dire, magari, che il mio riferimento non va oltre la foto? Ovviamente no: io non mi stavo riferendo a ciò che effettivamente ha causato l'apparenza della foto, né alla foto e basta, ma a ciò che in teoria avrebbe dovuto causarla; in questo caso, il mio tentativo di riferimento fallisce e basta.

Prendiamo il caso di un unicorno. Nessuno ha mai visto un unicorno, tutto ciò che noi possiamo pensare di un unicorno non è mai stato causato dall'osservazione di un unicorno. Supponiamo, allora, che ad un certo punto si scopra il fossile di un essere a forma di cavallo con un corno sulla fronte: dovremmo dire che abbiamo scoperto che esistono unicorni? Kripke, in effetti, risponde di no, ed è comprensibile visto che le nostre idee sugli unicorni, non essendo causate dagli unicorni, non hanno mai avuto una connessione fino a quel referente. Ma sembra invece che sia proprio legittimo dire che abbiamo scoperto che esistono unicorni perché quel fossile, posta l'intensione di unicorno, viene accettato nell'estensione. Nel caso di Swampman, similmente, quando dice "acqua" non c'è una catena causale che ci faccia risalire fino all'acqua; epperò ha comunque un'intensione tale che, quando vedrà l'acqua, sarà disposto a dire "acqua": se questo non vuol dire che aveva già un referente, vuol dire come minimo che il primo uso della parola farà da battesimo.

Spacca-cervello (letteralmente)

L'esempio di Swampman ci apre ad una serie di scenari enormi. Abbiamo già citato gli esperimenti mentali di Parfit (che si possono anche far risalire a Reid, il filosofo scozzese rispettoso del senso comune): se il mio cervello viene diviso in due e messo in due corpi, chi c'è ora? Ci sono io, ci sono due nuove persone, ci sono io e una nuova persona? Prima abbiamo cercato di rispondere dal punto di vista del concetto di identità, cioè in base al nostro modo di concettualizzare gli oggetti nel tempo. Ora, però, possiamo affrontare il problema da una prospettiva diversa, quella in prima persona.

Cosa vuol dire che io ho una prospettiva sul mondo? Be', io e l'amico siamo davanti ad uno specchio, e io vedo me dritto davanti e lui alla mia destra, e so che lui vede sé stesso dritto davanti e me alla sua sinistra, ma io non ho accesso diretto a questa sua esperienza che posso solo immaginare, e lui non ha accesso alla mia. Immaginiamo però che io sia stato soggetto ad un intervento, che il mio corpo calloso sia stato scisso, e che io ora abbia effettivamente accesso solo alle esperienze elaborate nel lobo destro, ho solo qualia associati al lobo destro. Potrei dire, allora, che il lobo sinistro è diventato inconscio come è inconscia l'attività del mio intestino, che solo il lobo destro supporta una visione in prima persona. Ma cosa dire, invece, se il lobo sinistro può pure supportare una visione in prima persona? Potrei immaginare che si tratta proprio di un'altra coscienza, distinta dalla mia tanto quanto la mia è distinta da quella dell'amico (con l'unica differenza che lui ha tutto un altro corpo esterno al mio); cioè, anche se io non ricevo qualia dal lobo sinistro, vi è "qualcuno" che ne fa esperienza.

Ora, come posso immaginare di ricevere qualia da solo uno dei due lobi, posso immaginare di estendere o diminuire la mia coscienza, riducendo il numero di qualia associati al cervello di cui faccio effettivamente esperienza, e posso immaginarmi così pienamente cosciente proprio come credo di essere ora (non credo, cioè, che il mio inconscio sia il conscio di qualcun altro, credo proprio che tutto ciò che sia esperibile nel mio cervello sia esperito da me), oppure posso immaginarmi minimamente cosciente, o totalmente inconsciente cioè inesistente. Ora, quando immagino di dividere il mio cervello tra due corpi, quando mi chiedo dove sono io e se ci sono

mi sto chiedendo, praticamente, a quali qualia continuo ad avere accesso (e se non vi sia invece qualcun altro che ha accesso ai qualia a cui avevo accesso io). Parlando d'identità, abbiamo accettato che non ci sia un fatto in merito, che se un certo corpo sia io oppure no dipende da quali sono gli altri individui che io personalmente considero presenti; questo, però, non si può ripetere in questo caso: che vi sia una certa prospettiva in prima persona sul mondo o che non vi sia è un fatto, non si può dire che è tutta e solo questione d'interpretazione.

Vi sono numerosi altri esempi sui quali è possibile riflettere, uno più complesso e spaventoso dell'altro.

Uno ispirato alla nave di Teseo potrebbe essere questo: vi è il mio cervello, e vi sono una serie di nanomacchine avanzatissime che hanno la capacità di assumere tutti i comportamenti caratteristici dei neuroni; ogni secondo, un mio neurone viene estratto e sostituito da una nanomacchina che inizia a comportarsi proprio come avrebbe fatto quel neurone; ovviamente io rimango io anche così, non basta certo perdere un neurone per distruggere la mia coscienza; dopo un po' di minuti, però, già metà del mio cervello è costituito da nanomacchine, e dopo ancora è stato totalmente sostituito; i miei neuroni, inoltre, non sono stati distrutti, ma sono stati conservati e ora vengono ricomposti secondo l'esatta struttura che avevano prima, ricreando il mio cervello. La domanda è: dove sono io? Supponiamo di vivere davvero un evento di questo tipo: che ne sarebbe della mia coscienza?

Analizzare questo problema può essere difficile perché vi sono molte variabili, per esempio la risposta può dipendere a seconda che si consideri la coscienza un fattore biologico (e allora le nanomacchine non possono essere coscienti) o funzionale (e allora possono esserlo tanto quanto un cervello di neuroni). Vediamo allora un esempio più semplice da analizzare.

Abbiamo due individui, A e B, cioè due prospettive sul mondo. I loro cervelli sono composti ovviamente da due lobi, che rappresentiamo come A-A e B-B:

A B
A-A B-B

Immaginiamo che esista un macchinario che è in grado di scambiare alla velocità della luce gli atomi o le cellule dei nostri corpi. Immaginiamo allora che questa macchina inizia a lavorare scambiando la materia dei nostri cervelli, e dopo poco otteniamo due nuovi organi:

50%
A? B?
A-B B-A

Dopo ancora, lo scambio è completo:

100%
? ?
B-B A-A

Ci chiediamo: se potessi avere accesso alle esperienze coscienti di A e B nel corso di questo esperimento, che cosa vedremmo? Possiamo rifletterci in tutti i modi, ma arriveremo probabilmente a quattro teorie possibili:

1. A (B) ha sempre e solo coscienza degli emisferi A (B), anche quando a metà processo si trovano uno in un cranio e uno in un altro
2. A (B) ha, al 50%, coscienza di A-B (B-A) e al 100% di B-B (A-A)
3. A (B) ha, al 50%, coscienza di A-B (B-A) e al 100% di A-A (B-B)
4. A e B scompaiono, i due cervelli ad un certo punto generano due nuovi individui C e D

La prima opzione discende dall'idea che una persona sia identificata strettamente dalla sua materia. È ovvio che si tratta di una tesi pazzza, anche se in alcuni casi potrebbe non suonare

così strana. Immaginiamo che durante una battaglia un cavaliere Sith venga tagliato perfettamente a metà da uno Jedi, un taglio che divide il corpo a metà seguendo scrupolosamente il piano sagittale com'è successo al Visconte Dimezzato, tanto che pochi secondi dopo una metà cade da una parte l'altra dall'altra: cosa ne è della coscienza del cavaliere? Il cervello è stato diviso a metà, anche tutta la spina dorsale è stata divisa a metà: è cosciente solo da un lato? Ha perso immediatamente coscienza per entrambi i lati? D'altro canto, se consideriamo che un uomo rimane cosciente per alcuni secondi anche dopo una decapitazione, non sembra un'immaginazione così strana l'idea che il cavaliere continui ad avere coscienza da entrambi i lati, come se ci fosse qualcosa al di là del corpo che garantisce per alcuni secondi una connessione, che continui a recepire informazioni da entrambi gli emisferi anche se lontani. Questa, però, è appunto solo un'immaginazione, utile per qualche anime cappa e spada, ma fin troppo difficile da credere: cosa mi impedirebbe, a questo punto, che non è possibile sostituire neanche un neurone del mio cervello perché non avrò mai coscienza se non dei neuroni che originalmente mi componevano? Ma questo è troppo controintuitivo, un neurone può ben essere sostituito.

La seconda e la terza opzione mostrano due diverse evoluzioni nel tempo. Nella seconda opzione ciò che identifica un certo soggetto come A è la sua "storia", il che vuol dire che A rimane A anche se tutta la materia del suo cervello viene scambiata con quella di B in quanto esiste una storia che porta dal cervello A-A al cervello B-B. Dico subito che, tra le prime tre, questa è l'opzione più convincente; epperò, ha i suoi grossi grossi problemi. Noi qui stiamo immaginando, infatti, un passaggio parziale di materia da un corpo all'altro, ed è facile individuare una storia. Immaginiamo però, allora, che i cervelli di A e B vengano smaterializzati e scambiati in modo istantaneo: in tal caso non è più così ovvio che sia questa la risposta, non esiste più una storia che porta da uno stato all'altro, eppure il semplice aumentare la velocità del processo, o la quantità di materia spostata per volta, non dovrebbe essere un fattore rilevante. Difendere questa posizione richiederebbe, in primo luogo, di trovare una nuova difesa per l'essenzialità dell'origine e, in secondo luogo, di essere in grado di spiegare esattamente in quali casi si possa affermare che la "storia causale" di un corpo continua e in quali casi invece viene spezzata (perché se sposto un atomo per volta mi pare continuare, ma se sposto un emisfero in modo istantaneo invece no?).

Nella terza opzione, invece, l'individuo A rimane associato ad un certo cervello finché non si supera una certa soglia, per esempio finché il 50%+1 dei neuroni non viene spostato nel cranio di B, in quel caso anche la sua coscienza si sposta nell'altro corpo. Questo approccio può essere accettabile nella discussione dell'identità corporale, ma dal punto di vista dell'esperienza è troppo strana, è un puro problema delle sorite ed è impossibile individuare una soglia oggettiva, che pure dovrebbe esserci se fosse questo il meccanismo.

La quarta opzione, per esclusione, risulta attraente ed è vicina all'ultima proposta fatta discutendo l'identità; ma mentre per l'identità può essere accettabile, per la coscienza non è così semplice: posso ben dire che non ci sia un fatto in merito alla mia identità, ma mi sembrerebbe assurdo rie che non c'è un fatto in merito al mio essere cosciente di certe cose e incosciente di certe altre, di essere cosciente di certi processi del mio cervello ed essere totalmente incosciente di tutti i processi che avvengono nel cervello dell'amico. Supponiamo che la proposta sia corretta, che ad un certo punto io cesso di essere cosciente di qualsiasi informazione, ma che l'informazione di quel cervello continua a supportare una prospettiva in prima persona (solo non quella che c'era all'inizio): questo comunque rimane un problema delle sorite, non è chiaro quando dovrebbe accadere esattamente, non è chiaro come vedrei evolversi la mia coscienza del tempo se io fossi il cervello A-A.

Tutte le opzioni paiono insoddisfacenti, ed è probabilmente inutile continuare a cercare di arrivare a qualche conclusione in modo puramente teorico: la questione, infatti, è empirica, è un qualcosa che io potrei vedere se fossi sottoposto ad un'operazione del genere, se io fossi il cervello A-A potrei vedere, tramite l'evolversi della mia esperienza, qual è l'opzione corretta, anche se con tutta probabilità non sarei in grado di comunicarla a nessuno (non comunicerei che ciò che il mio cervello può comunicare, se il mio cervello fosse strutturato in modo tale da dire "tutto bene, non è cambiato niente" direi così anche se in realtà avessi l'impressione di grandi stravolgimenti). D'altro canto, potrei anche non riuscire davvero a conoscere cosa accade nel senso che potrei morire nel processo, oppure potrei semplicemente essere determinato a credere certe cose, a pensare che realmente non sia cambiato nulla.

Dove sta il problema? Quando ci chiediamo "cosa accade nel fenomeno" in seguito a certi cambiamenti "nel noumeno", stiamo facendo un'operazione dubbia. Ci stiamo chiedendo, posto un certo evento descritto in terza persona (una certa modifica nel cervello), a che evento in prima persona corrisponde (che esperienza). Come ci chiediamo "com'è essere un pipistrello", qui ci chiediamo "com'è essere un cervello sottoposto a queste operazioni". Il punto è che la terza persona e la prima persona paiono troppo diverse per essere relazionate in modo così diretto e semplice. Il mondo osservato in terza persona è discontinuo, fatto di parti, di particelle, di istanti; il mondo osservato in prima persona, invece, è continuo e unito, le sostanze appaiono continue e infinitamente divisibili, l'esperienza è un tutt'uno unico senza componenti (un libro blu è un libro blu, non è l'esperienza di un libro più l'esperienza del blu separatamente). Nel mondo in terza persona, se un feto è una chimera, nato dalla fusione di due zigoti con DNA diversi, posso ben dire che il suo corpo è "composto dal 50% dal corpo A e al 50% dal corpo B", ma sarei assurdo a dire in prima persona che la sua mente è al 50% la mente A e al 50% la mente B, la mente ha settori ma non componenti, non ci sono "cellule di mente". Sembra che il discorso non sia affrontabile se non dividendo le due prospettive. È come se la prima e la terza persona fossero due lingue diverse, possiamo immaginare che esistano proposizioni, potremmo dire cioè che la frase in terza "evento neurale x" e la frase in prima "dolore" esprimono la stessa identica cosa; ma, come sappiamo, la traduzione è indeterminata, è impossibile arrivare a far corrispondere i due insiemi di frasi in modo oggettivo, come vi sono infiniti modi possibili e legittimi di collegare "gavagai" ad un certo significato così vi saranno infiniti modi di collegare certi eventi neurali a certe esperienze mentali, di modo che possono risultare infiniti diversi ritratti dell'io.

Non dovremmo dire, allora, che non accade se non proprio ciò che sembra accadere: le cellule si spostano in un certo modo, e non ha senso chiedersi come appaia questo evento dalla prima persona. Se si vuole parlare della prima persona, be', bisognerebbe ridescrivere il problema tramite termini della prima persona, senza iniziare da un evento che riguarda i cervelli "fuori di me"; ne consegue, comunque, che il problema non può ricevere risposta se non vivendo quell'evento.

Una proposta d'idealismo

A ben pensarci, però, si nota che la prima persona ha priorità sulla terza. Sembra, cioè, che ciò che osserviamo in terza persona generi la prima; di fatto, però, noi non osserviamo il mondo che in prima persona.

La nostra intuizione è che vi siano diversi soggetti, vi siano cioè più prospettive in prima persona, ma che tutti osservino lo stesso mondo, che esista un unico mondo in terza persona. Come notano diversi filosofi, tra i quali Sellars e Dummett, l'ambizione della scienza è quella di dare una descrizione in pura terza persona del mondo, cioè ha l'ambizione di andare a descrivere quel mondo che è comune a tutte le esperienze in quanto fa da base ad esse, vuole

descrivere il mondo così com'è in modo indipendente da qualsiasi prospettiva in prima persona. Io vedo me stesso in prima persona e l'amico in terza, l'amico vede sé stesso in prima persona e me in terza, la scienza mette insieme le due prospettive di terza persona per creare una descrizione oggettiva del mondo di sicuro successo empirico in quanto permette di prevedere cosa ognuno, pur avendo una diversa prospettiva sul mondo, vedrà in vari casi.

Immaginiamo allora di nuovo così... Io e l'amico siamo davanti allo specchio, uno di fianco all'altro; questa volta, però, abbiamo anche la testa scoperchiata, in modo tale da vedere i nostri cervelli. Io posso toccarmi un punto del mio cervello, e noto che, oltre a vedere il mio dito che tocca il mio cervello, ho in più anche un'ulteriore esperienza, diciamo che vedo rosso perché ho toccato una qualche zona adibita all'analisi del colore. Quando l'amico fa lo stesso, e tocca una zona del suo cervello, io non ho l'esperienza del rosso, vedo solo lui che tocca, e lo sento dire che anche lui ha avuto esperienza del rosso. Questo vuol dire che all'evento neurale W (dovuto al toccare il cervello) segue una esperienza X (il vedere rosso), mentre all'evento neurale Y (dovuto al guardare l'amico che si tocca il cervello) segue l'esperienza Z (il vederlo fare ciò, invece del vedere rosso).

Da questo esempio, io posso dedurre che io e lui abbiamo diverse prospettive sul mondo e che gli eventi cerebrali correlano con eventi mentali, deduco che io ho visto rosso toccando il cervello perché io sono quel cervello, mentre non ho visto rosso quando l'ha toccato lui perché io non sono quello, epperò penso che lui abbia comunque avuto quell'esperienza a me inaccessibile. La prospettiva, nello specifico, è resa mostrando come il toccarsi il proprio cervello (evento W) e l'osservare un altro toccarsi il proprio (evento Y) sono stati diversi ai quali seguono esperienze diverse (insomma, che il cervello A sia nello stato W, e che il cervello A osservi il cervello B essere nello stato W - cioè che il cervello A sia nello stato Y - sono due esperienze diverse), per questo in un caso mi pare che l'evento neurale implichi un'esperienza soggettiva mentre lo stesso evento in un altro cervello non lo fa.

Ora, l'idea della scienza "in terza persona" è che l'evento osservato in X (la visione del rosso) non sia una corretta interpretazione di ciò che sta accadendo, ciò che sta accadendo è quanto io posso osservare in Z (vedo un altro cervello avere l'evento W). Questa idea, così descritta, suona ben strana. Io vedo l'evento W come l'esperienza X, la scienza però mi dice che se voglio conoscere W io non devo guardare X, ma bensì Z, causata da Y; cioè, io non posso capire l'evento W se non tramite l'evento Y, per capire un evento non devo viverlo ma bensì devo vederlo vivere ad altri. Questo ci fa cadere ovviamente in un regresso infinito: per capire lo stato neurale W devo essere nello stato Z, allora per capire Z dovrò essere nello stato beta, e per capirlo dovrò essere nello stato gamma, e via così. Ma soprattutto: perché mai dovrei pensare che l'esperienza X sia ciò che appare nell'esperienza Z (cioè, che il vedere rosso sia un certo evento neurale), anziché essere proprio ciò che appare essere (o anche essere qualcos'altro di totalmente diverso da tutte le nostre esperienze possibili)?

Il punto è che, come dice Sellars, "non esiste un punto di vista da-nessuna-parte", non è possibile descrivere un mondo che nessuno vede, il mondo può essere descritto solo dalla prospettiva "di qualcuno". Come voleva Kant, possiamo parlare solo del fenomeno, possiamo parlare del mondo solo così com'è una volta concettualizzato, non possiamo parlare del mondo a prescindere da qualsiasi concetto, è un'insensatezza.

Di fatto, non è la terza persona che causa la prima, di fatto io osservo il mondo in prima persona e solo dopo deduco l'esistenza di un mondo in terza persona (forse, ontologicamente, la terza persona causa la prima, ma epistemologicamente la terza persona discende dalla prima). Se consideriamo tutti i problemi che il dualismo e il monismo neutrale hanno, insieme ai problemi appena descritti che sorgono quando si cerca di far emergere un evento mentale da uno neurale, risulta evidente che l'unica alternativa al fisicalismo è una sorta di idealismo o, se

il termine fa vergognare i filosofi contemporanei, una forma di "fenomenalismo", o "sensualismo": invece di cercare di fondere le descrizioni in terza persona, il modo in cui tutti i soggetti vedono gli altri, mettiamo insieme quelle in prima, il modo in cui ognuno vede sé stesso, e mettiamo il mondo esterno tra parentesi in quanto non ha senso parlarne a prescindere dal fenomeno.

Se faccio così, ne risulterà una descrizione di questo tipo: all'esperienza W (il vedermi toccare il cervello) segue l'esperienza X (il vedere rosso), mentre all'esperienza Y (il vedere l'amico toccarsi il cervello) segue l'esperienza Z (l'amico che dice di vedere rosso). Non si descrive il modo in cui gli eventi fisici causano i mentali, bensì tutto ciò che si può descrivere è solo il modo in cui l'esperienza si evolve: non c'è alcuna differenza tra le cose come sono e le cose come appaiono.

Mi rendo conto che ci si sta spostando in cattive acqua. Si sta derivando una sorta di idealismo dai limiti dell'empirismo, proprio come aveva fatto il vescovo Berkeley il quale credeva che tutto ciò che vediamo non fosse che l'illusione creata dalla luce proveniente da Dio. Non è però un idealismo soggettivo come il suo, è comunque un idealismo oggettivo al modo di Peirce e William Hocking; l'ideale, però, è forse l'idealismo di Schelling perché mantiene una teoria corrispondentista della realtà. L'ideale, infatti, è l'accettare un "realismo diretto" per il quale la realtà è proprio così come la percepiamo, è essenzialmente coscienza; questa realtà, però, deve poter essere interpretata e strutturata dal nostro pensiero, e la nostra conoscenza d'essa deve rimanere fallibile per non cadere nel solipsismo (bisogna mantenere la possibilità di parlare di oggetti comuni, di errare, di osservarci a vicenda, di cambiare nel tempo...): in questo modo la posizione non diventa revisionista, cioè è possibile continuare a ritenere perfettamente valido tutto ciò che la scienza ci dice sul mondo, con l'unica differenza che non ci informa su qualcosa che sta dietro l'esperienza, ma dentro l'esperienza. Più che idealismo è una "fenomenologia critica" kantiana.

Vi è qualche filosofo contemporaneo che ha preso sul serio queste idee: abbiamo il rivale di Wittgenstein A. C. Ewing, il fenomenalista Brian Jonathan Foster, il teologo Paul Moser, il continentale Vittorio Hosle, ma il più interessante è Bernardo Kastrup, un filosofo con due dottorati, uno in filosofia appunto e uno in informatica (ha lavorato anche al CERN), che scrive lungamente su come il mondo sia sostanzialmente immaginato, e come la coscienza stia ad un piano più fondamentale della materia stessa. Forzando un po', si può trovare qualche parallelo con il filosofo della mente italiano Riccardo Manzotti, anche lui come Kastrup specializzato in informatica e con un dottorato in robotica: come Tommaso d'Aquino combatteva lo scetticismo mostrando una mela e dicendo "questa è una mela, chi non è d'accordo se ne vada pure", anche Manzotti gira sempre con una mela, e quando si chiede dov'è l'esperienza della mela, dov'è la mente, dice che non è nella testa, ma è proprio lì dov'è la mela, l'aver coscienza della mela è la mela, non c'è differenza tra sé e il mondo; non voglio però certo spingermi a dire che è un fenomenalista e affibbiargli idee che non sostiene.

Un problema evidente di tale approccio sta nello spiegare com'è possibile condividere un mondo. Io sto camminando per un sentiero di montagna, e compare un leone. È un'allucinazione, chiaramente: ma perché? Perché lì non c'è un leone, viene da dire: ma in un approccio fenomenalista non posso dirlo, non posso appellarmi ad una non-corrispondenza con il mondo esterno per distinguere le allucinazioni da altri tipi di percezione. Allora potrei dire perché lo vedo solo io, nessun altro vede quel leone: ma, anche se non è solipsismo, non posso dirlo, io posso dire al massimo che percepisco gli altri dire di non vederlo, ma in fondo ci sono un sacco di cose, reali, che non tutti vediamo, per esempio non tutti siamo in grado di distinguere un animale mimetizzato tra le foglie, e il fatto che io lo veda e gli altri mi dicono di non vederlo non rende certo quell'animale meno reale (poco conta il fatto che, nel caso del

leone, questo non mi pare certo cercare di mimetizzarsi). È molto difficile rispondere a questa domanda, bisogna trovare il modo, se si vuole difendere la sensatezza della posizione, di sviluppare un "disgiuntivismo" idealista, una teoria che sia in grado di disgiungere le allucinazioni dalle percezioni "reali".

(si può fare così: di norma, l'oggetto reale causa sense-data mentali ma inconsci (le modificazioni fisiche nel cervello causate dalla luce) che poi vengono analizzate. Poiché, in caso di allucinazione, non c'è l'oggetto, non ci sono neanche i sense data. Ora, nell'idealismo, gli oggetti esterni non ci sono, ma i sense data ideali potrebbero esservi comunque, e potrebbero essere assenti in caso di allucinazione)

Bibliografia

- Aczel, A. D. (2015). *Il Mistero dell'Alef. La Ricerca dell'Infinito tra Matematica e Misticismo*. Milano: Il Saggiatore.
- Adams, R. M. (1997). A Modified Divine Command Theory of Ethical Wrongness. In T. L. Carson, & P. K. Moser, *Morality and the Good Life*. Oxford: Oxford University Press.
- Adams, R. M. (1987). *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford: Oxford University Press.
- Adorno, T. (2015). *Minima Moralia. Meditazioni della Vita Offesa*. Torino: Einaudi.
- Agostino. (2013). *Confessioni*. Roma: Newton Compton.
- Agostino. (2015). *La Città di Dio*. Milano: Mondadori.
- Agostino. (2012). *La Trinità*. Milano: Bompiani.
- Albert, H. (1968). *Treatise on Critical Reason*. Princeton, New Jersey: Princeton University Press.
- Alberto-Magno. (2007). *L'unità dell'Intelletto*. Milano: Bompiani.
- Althusser, L., Balibar, É., Establet, R., Rancière, J., & Macherey, P. (2006). *Leggere il Capitale*. Sesto San Giovanni: Mimesis.
- Anscombe, G. E. (2016). *Intenzione*. Roma: Edusc Santa Croce.
- Anscombe, G. E. (1958). Modern Moral Philosophy. *Philosophy vol.33 (124)* , 1-19.
- Antognazza, M. R. (2015). *Leibniz - Una Biografia Intellettuale*. Milano: Hoepli.
- Aristotele. (1999). *Etica Nicomachea*. Roma-Bari: Laterza.
- Aristotele. (2010). *La Metafisica*. Torino: UTET.
- Aristotele. (2003). *Organon*. Milano: Adelphi.
- Austin, J. L. (1975). *How to Do Things With Words*. Cambridge, Massachusetts: Harvard University Press.
- Austin, J. L. (1950). Truth. *Aristotelian Society Supplementary Volume (24: 1)* , 111-172.
- Averroè. (1994). *Il Trattato Decisivo sulla Connessione della Religione con la Filosofia*. Milano: Rizzoli.
- Ayer, A. J. (1952). *Language, Truth and Logic*. New York City: Dover Publications.
- Ayer, A. J. (1959). *Logical Positivism*. Glencoe, Scotland: Free Press.
- Bach, K. (1985). A Rationale for Reliabilism. *The Monist 68 (2)* , 246-263.
- Bachelard, G. (2018). *Il Nuovo Spirito Scientifico*. Sesto San Giovanni: Mimesis.
- Bachelard, G. (2006). *La Poetica dello Spazio*. Bari: Dedalo.

- Bachelard, G. (2020). *L'Attività Razionalistica della Fisica Contemporanea*. Milano: Jaca Book.
- Bachelard, G. (2015). *Psicanalisi delle Acque*. Milano: Red Edizioni.
- Bachelard, G. (2016). *Saggio Sulla Conoscenza Approssimata*. Sesto San Giovanni: Mimesis.
- Bagaric, M., & Clarke, J. (2007). *Torture: When the Unthinkable is Permissible*. Albany, New York: State University of New York Press.
- Baron, M. (2003). Acting From Duty. In I. Kant, *Groundwork for the Metaphysics of Morals* (p. 98-99). New Haven: Yale University Press.
- Barrow, J. D. (2006). *L'Infinito. Breve Guida ai Confini dello Spazio e del Tempo*. Milano: Mondadori.
- Barthes, R. (2003). *Il Grado Zero della Scrittura*. Torino: Einaudi.
- Barthes, R. (2002). *Ronald Barthes*. Torino: Einaudi.
- Barthes, R. (1981). *S/Z*. Torino: Einaudi.
- Barwise, J., & Perry, J. (1981). Semantic innocence and uncompromising situations. *Midwest Studies in Philosophy* 6 (1) , 387-404.
- Baudrillard, J. (2008). *Simulacri e Impostura. Bestie, Beaubourg, Apparenze e Altri Oggetti*. Milano: PGreco.
- Beccaria, C. (1764). *Dei Delitti e delle Pene*.
- Benjamin, W. (2011). *L'Opera d'Arte nell'Epoca della sua Riproducibilità Tecnica*. Torino: Einaudi.
- Bentham, J. (2020). *Introduzione ai Principi della Morale e della Legislazione*. Padova: Primiceri.
- Bentham, J. (1997). *Panopticon, Ovvero la Casa d'Ispezione*. Venezia: Marsilio.
- Bentham, J. (1830). *The Rationale of Reward*. London: Robert Heward.
- Bentham, J. (2004). *The Theory of Legislation*. London: Continuum International.
- Berdjaev, N. (2017). *Nuovo Medioevo*. Roma: Fazi.
- Berger, P. L., & Berger, B. (1977). *Sociologia. La Dimensione Sociale della Vita Quotidiana*. Bologna: Il Mulino.
- Bergson, H. (1907). *L'Evoluzione Creatrice*.
- Berkeley, G. (1710). *Trattato sui Principi della Conoscenza Umana*.
- Berlin, I. (1998). *Il Riccio e la Volpe e Altri Saggi*. Milano: Adelphi.
- Berlin, I. (2001). *Le Radici del Romanticismo*. Milano: Adelphi.
- Berlin, I. (1959). *Tolstoj e la Storia*. Milano: Lerici.
- Berlin, I. (1994). *Il Legno Storto dell'Umanità. Capitoli della Storia delle Idee*. Milano: Adelphi.

- Berruto, G., & Cerruti, M. (2011). *La Linguistica. Un Corso Introduttivo*. Torino: UTET.
- Berto, F. (2010). *L'Esistenza non è Logica. Dal Quadrato Rotondo ai Mondi Impossibili*. Roma-Bari: Laterza.
- Berto, F. (2008). *Logica da Zero a Gödel*. Roma-Bari: Laterza.
- Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Blackmon, J. (2005). *Disparities of Meaning for a Solution to the Liar (articolo non pubblicato)*. Tratto il giorno febbraio 2020, 14 da Researchgate: https://www.researchgate.net/publication/299362087_Disparities_of_Meaning_for_a_Solution_to_the_Liar
- Bloch, E. (2019). *Il Principio Speranza, Vol. 1-3*. Sesto San Giovanni: Mimesis.
- Boezio. (2014). *La Consolazione della Filosofia*. Torino: UTET.
- Bolzano, B. (1837). *Dottrina della Scienza*.
- Bonjour, L. (2002). *Epistemology: Classic Problems and Contemporary Responses*. Lanham, Maryland: Rowman & Littlefield.
- Bonjour, L. (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy* 5, 53-73.
- Bonjour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, Massachusetts: Harvard University Press.
- Bourget, D., & Chalmers, D. J. (2014). What Do Philosophers Believe? *Philosophical Studies* 170 (3), 465-500.
- Boyd, R. (1988). How to Be a Moral Realist. In G. Sayre-McCord, *Essays on Moral Realism* (p. 187-228). Ithaca, New York: Cornell University Press.
- Brandom, R. (2019). *A Spirit of Trust: A Reading of Hegel's Phenomenology*. Cambridge, Massachusetts: Harvard University Press.
- Brandt, R. (1959). *Ethical Theory*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Brentano, F. (1874). *La Psicologia dal Punto di Vista Empirico*.
- Bridges, J. (2014). Rule-Following Skepticism, Properly So Called. In C. James, & A. Kern, *Varieties of Skepticism*. Berlin: DeGruyter.
- Brink, D. (2001). Realism, Naturalism, and Moral Semantics. *Social Philosophy and Policy*, 18, 154-176.
- Bronner, B., & Goldstein, S. (2018). A Stronger Doctrine of Double Effect. *Australasian Journal of Philosophy*, 96 (4), 793-805.
- Brosnan, K. (2011). Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge? *Biology and Philosophy*, 26, 51-64.

- Brouwer, L. E. (2015). *Vita, Arte e Mistica*. Milano: Adelphi.
- Bruno, G. (1985). *Spaccio de la Bestia Trionfante*. Milano: Rizzoli.
- Buber, M. (2011). *Il Principio Dialogico e Altri Saggi*. Cinisello Balsamo: San Paolo.
- Butler, J. (2014). *Fare e Disfare il Genere*. Sesto San Giovanni: Mimesis.
- Butler, J. (2017). *Questioni di Genere. Il Femminismo e la sovversione dell'identità*. Roma-Bari: Laterza.
- Button, T. (2013). *The Limits of Realism*. Oxford: Oxford University Press.
- Calderoni, M. (2007). *Scritti sul Pragmatismo*. Acireale: Bonanno.
- Cambridge, L. L. (2005). *David Edmonds; John Eidinow*. Milano: Garzanti.
- Campanella, T. (1997). *La Città del Sole*. Roma-Bari: Laterza.
- Campbell, R. (2019). Moral Epistemology. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)* (p. <https://plato.stanford.edu/archives/win2019/entries/moral-epistemology/>).
- Camus, A. (1942). *Lo Straniero*.
- Cantor, G. (2020). *La Formazione della Teoria degli Insiemi (scritti 1872-1899)*. Sesto San Giovanni: Mimesis.
- Carnap, R. (1996). The Elimination of Metaphysics Through Logical Analysis of Language. In S. Sarkar, *Logical Empiricism at its Peak: Schlick, Carnap, and Neurath* (p. 10-31). New York City: Garland.
- Cartesio. (1637). *Discorso sul Metodo*.
- Cartesio. (1641). *Meditazioni Metafisiche*.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71 (5) , 805-819.
- Cassirer, E. (2018). *Il Concetto di Sostanza e il Concetto di Funzione*. Brescia: Morcelliana.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, D., Manley, D., & Wesserman, R. (2009). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.
- Chomsky, N. (1975). *La Grammatica Trasformativale*. Torino: Bollati Boringhieri.
- Chomsky, N., & Foucault, M. (2005). *Della Natura Umana. Invariante Biologico e Potere Politico*. Roma: DeriveApprodi.
- Church, A. (1943). Review of Rudolf Carnap, Introduction to Semantics. *The Philosophical Review* (52) , 298-304.

- Churchland, P. (1992). *La Natura della Mente e la Struttura della Scienza. Una Prospettiva Neurocomputazionale*. Bologna: Il Mulino.
- Cicerone, M. T. (2019). *De Officiis. Quel che è Giusto Fare*. Milano: Einaudi.
- Cicerone, M. T. (2019). *Le Tuscolane*. Siena: Liberamente.
- Ciraci, F. (2006). *Verso l'Assoluto Nulla. La Filosofia della Redenzione di Philipp Mainländer*. Lecce: Pensa Multimedia.
- Cohen, S. M., & Reeve, C. D. (2020). Aristotle's Metaphysics. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)* (p. <https://plato.stanford.edu/archives/win2020/entries/aristotle-metaphysics/>).
- Collingwood, R. G. (1971). *Il Nuovo Leviatano o Uomo, Società, Civiltà, Barbarie*. Milano: Giuffrè.
- Collins, R. (1980). *Sociologia*. Bologna: Zanichelli.
- Comesaña, J. (2014). There is no Immediate Justification. In M. Steup, & E. (. Sosa, *Contemporary Debates in Epistemology*. Malden, Massachusetts: Blackwell.
- Comte, A. (2017). *Discorso sullo Spirito Positivo*. Milano: Hachette.
- Confucio. (2006). *Dialoghi*. Milano: Feltrinelli.
- Copernico, N. (1979). *Opere*. Torino: UTET.
- Copp, D. (2001). Realist-Expressivism: A Neglected Option for Moral Realism. *Social Philosophy and Policy*, 18 , 1-43.
- Coscolluela, V. (1995). *The Ethics of Suicide*. New York City: Garland.
- Cowie, C. (2018). Companions in Guilt Arguments. *Philosophy Compass* 13 (11) , e12528.
- Cox, C. (1926). *Genetic Studies of Geniuses, 2 vol*. Redwood City, California: Stanford University Press.
- Croce, B. (1909). *Logica come Scienza del Concetto Puro*.
- Csikszentmihályi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York City: Harper & Row.
- Cusano, N. (2017). *Opere Filosofiche, Teologiche e Matematiche*. Milano: Bompiani.
- Damasio, A. (1995). *L'Errore di Cartesio. Emozione, Ragione e Cervello Umano*. Milano: Adelphi.
- Dancy, J. (1995). In Defence of Thick Concepts. *Midwest Studies in Philosophy*, 20 , 263-279.
- d'Aosta, A. (2002). *Monologio e Proslogio*. Milano: Bompiani.
- d'Aquino, T. (2016). *Compendio di Teologia*. Torino: UTET.
- d'Aquino, T. (1270). *De Aeternitate Mundi*.

d'Aquino, T. (1995-2013). *La Somma Teologica in 35 Volumi*. Bologna: Edizioni Studio Domenicano.

Davidson, D. (2005). A Nice Derangement of Epitaphs. In D. Davidson, *Truth, Language, and History Philosophical Essays Volume 5*. Oxford: Oxford University Press.

Davidson, D. (2003). *Soggettivo, Intersoggettivo, Oggettivo*. Milano: Raffaello Cortina.

Davidson, D. (1969). True to the Facts. *Journal of Philosophy* 66 (21) , 748-764.

Davidson, D. (1994). *Verità e Interpretazione*. Bologna: Il Mulino.

Davidson, D. (1994). *Verità e Interpretazione*. Bologna: Il Mulino.

Davis, M., & Putnam, H. (1960). A Computing Procedure for Quantification Theory. *Journal of the ACM* 7 (3) , 201-215.

Dawkins, R. (1995). *Il Gene Egoista. La Parte Immortale di ogni Essere Vivente*. Milano: Mondadori.

Dawkins, R. (2007). *L'Illusione di Dio*. Milano: Mondadori.

de Beauvoir, S. (1949). *Il Secondo Sesso*.

de la Barca, P. C. (1635). *La Vida es Sueño*.

de Montaigne, M. (2014). *Saggi*. Milano: Bompiani.

De Saussure, F. (1916). *Corso di Linguistica Generale*.

Deleuze, G., & Guattari, F. (2002). *L'Anti-Edipo. Capitalismo e Schizofrenia*. Torino: Einaudi.

della Mirandola, G. P. (2010). *Dell'Ente e dell'Uno*. Milano: Bompiani.

Demostene. (1987). *Prima Olintiaca*. Roma: Avia Pervia.

Dennett, D. (2015). *L'Idea Pericolosa di Darwin. L'Evoluzione e i Significati della Vita*. Torino: Bollati Boringhieri.

Derby, J. W. *A Philosopher by Lamplight*. Derby Museum and Art Gallery, Derby.

Derrida, J. (1967). *Della Grammatologia*.

Derrida, J. (2002). *La Scrittura e la Differenza*. Torino: Einaudi.

Dewey, J. (2014). *Esperienza e Educazione*. Milano: Raffaello Cortina Editore.

Di Nucci, E. (2014). Eight Arguments Against Double Effect. *Proceedings of the XXIII Kongress der Deutschen Gesellschaft für Philosophie* .

di Ockham, G. (2015). *Dialogo sul Papa Eretico*. Milano: Bompiani.

Diels, H. A., & Kranz, W. (1903). *I Presocratici*.

Dilthey, W. (2007). *Introduzione alle Scienze dello Spirito*. Milano: Bompiani.

- Dostoevskij, F. (1866). *Delitto e Castigo*.
- Dreier, J. (2006). Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution. In R. Shafer-Landau, *Oxford Studies in Metaethics* (p. 217-234). Oxford: Oxford University Press.
- Dreier, J. (1993). The Supervenience Argument against Moral Realism. *Southern Journal of Philosophy*, 30 , 13-38.
- Dretske, F. (1971). Conclusive Reasons. *Australasian Journal of Philosophy* 49 (1) , 1-22.
- Dretske, F. (1970). Epistemic Operators. *Journal of Philosophy* 67 , 1007-1023.
- Dretske, F. (2005). Is Knowledge Closed Under Known Entailment? In M. Steup, & E. (. Sosa, *Contemporary Debates in Epistemology*. Malden, Massachusetts: Blackwell.
- Dretske, F. (2000). *Perception, Knowledge and Belief*. Cambridge: Cambridge University Press.
- Dretske, F. (2003). Skepticism: What Perception Teaches. In S. Luper, *The Sceptics* (p. 105-118). Hampshire: Ashgate.
- Dreyfus, H., & Dreyfus, S. (1986). *Mind Over Machine*. Oxford: Basil Blackwell.
- Dummett, M. (1986). *La Verità e Altri Enigmi*. Milano: Il Saggiatore.
- Dummett, M. (2001). *Origini della Filosofia Analitica*. Torino: Einaudi.
- Dummett, M. (2008). *Pensiero e Realtà*. Bologna: Il Mulino.
- Dummett, M. (2006). *Verità e Passato*. Milano: Raffaello Cortina.
- Dummett, M., & Flew, A. (1954). Can and Effect Precede its Cause? *Aristotelian Society Supplementary Volume*, 28 (1) , 27-62.
- Edmonds, D. (2014). *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton, New Jersey: Princeton University Press.
- Eger, R. J., & Maridal, J. H. (2015). A Statistical Meta-analysis of The Wellbeing Literature. *International Journal of Wellbeing*, 5 (2) , 45-74.
- Einstein, A. (2002). *The Berlin Years: Writings, 1918-1921*. Princeton, New Jersey: Princeton University Press.
- Elliot, R., & Jamieson, D. (2009). Progressive Consequentialism. *Philosophical Perspectives*, 23 , 241-251.
- Emerson, R. W. (2005). *Diventa Chi Sei*. Roma: Donzelli.
- Emerson, R. W. (2009). *Uomini Rappresentativi*. L'Aquila: REA.
- Epicuro. (2007). *Opere. Frammenti. Testimonianze sulla Vita*. Roma-Bari: Laterza.
- Epitteto. (1996). *Manuale*. Milano: Rizzoli.
- Erasmus. (2005). *Elogio della Follia*. Torino: Einaudi.

- Eriugena, G. S. (2013). *Divisione della Natura*. Milano: Bompiani.
- Fabro, C. (2006). *Fenomenologia della Percezione*. Segni: Istituto del Verbo Incarnato.
- Fabro, C. (2010). *L'Assoluto nell'Esistenzialismo*. Segni: Istituto del Verbo Incarnato.
- Fabro, C. (2005). *Neotomismo e Suarezismo*. Segni: Istituto del Verbo Incarnato.
- Falcon, A. (2019). Aristotle on Causality. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)* (p. <https://plato.stanford.edu/archives/spr2019/entries/aristotle-causality/>).
- Falk, W. D. (1947). "Ought" and Motivation. *Proceedings of the Aristotelian Society*, 48 , 492-510.
- Ferraris, M. (2014). *Manifesto del Nuovo Realismo*. Roma-Bari: Laterza.
- Feuerbach, L. (1841). *L'Essenza del Cristianesimo*.
- Feyerabend, P. K. (2013). *Contro il Metodo*. Milano: Feltrinelli.
- Fichte, J. G. (1794). *Fondamenti dell'Intera Dottrina della Scienza*.
- Fichte, J. (1992). *La Missione del Dotto*. Milano: Mursia.
- Ficino, M. (2013). *Sopra lo Amore ovvero Convito di Platone*. Milano: SE.
- Firth, R. (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research*, 12 (3) , 317-345.
- Fitch, F. (1963). Logical Analysis of some Value Concepts. *The Journal of Symbolic Logic* (28) , 135-142.
- Flek, L. (1981). *Genesis and Development of a Scientific Fact*. Chicago, Illinois: The University of Chicago Press.
- Fletcher, J. (1997). *Situation Ethics: The New Morality*. Louisville, Kentucky: Westminster John Knox Press.
- Foot, P. (1967). The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, 5 , 5-15.
- Foot, P. (2008). *Virtù e Vizi*. Bologna: Il Mulino.
- Ford, P. L. (1921). *The Many-sided Franklin*. New York City: Century Company.
- Foss, B. (. (1968). *I Nuovi Orizzonti della Psicologia*. Torino: Bollati Boringhieri.
- Foucault, M. (1998). *Nascta della Clinica. Una Archelogia dello Sguardo Medico*. Torino: Einaudi.
- Foucault, M. (2014). *Sorvegliare e Punire*. Torino: Einaudi.
- Foucault, M. (2013-2019). *Storia della Sessualità, vol. 1-4*. Feltrinelli.
- Fraassen, B. V. (1984). Belief and the Will. *Journal of Philosophy* 81 (5) , 235-256.

- Frege, G. (2021). *L'Alfabeto del Pensiero*. Roma: Castelvecchi.
- Frege, G. (2019). *Logica, Pensiero e Linguaggio. I Fondamenti dell'Aritmetica e Altri Scritti*. Roma-Bari: Laterza.
- Frege, G. (2007). *Senso, Funzione e Concetto. Scritti Filosofici 1891-1897*. Roma-Bari: Laterza.
- Fries, J. F. (1807). *Neue oder anthropologische Kritik der Vernunft [Nuova Critica Antropologica della Ragione]*.
- Fromm, E. (2018). *Avere o Essere?* Milano: Mondadori.
- Gadamer, H.-G. (1960). *Verità e Metodo*.
- Galilei, G. (1632). *Dialogo Sopra i Due Massimi Sistemi del Mondo*.
- Garrigou-Lagrange, R. (2012). *Essenza e Attualità del Tomismo*. Verona: Fede & Cultura.
- Garrigou-Lagrange, R. (2015). *La Sintesi Tomistica*. Verona: Fede & Cultura.
- Geach, P. (1965). Assertion. *Philosophical Review*, 74 , 449-465.
- Gentile, G. (1916). *Teoria Generale dello Spirito come Atto Puro*.
- George, L. K., Larson, D. B., Koenig, H. G., & McCullough, M. E. (2000). Spirituality and Health: What We Know, What We Need to Know. *Journal of Social and Clinical Psychology*, 19 (1) , 102-116.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis* (23) , 121-123.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Gilman, S. J. (2012). The Use of Anencephalic Infants as an Organ Source: an On-Going Question. *Elon Law Review*, 4 (1) , 71-92.
- Gilson, É. (1984). *Dio e la Filosofia*. Milano: Massimo.
- Gödel, K. (2006). *La Prova Matematica dell'Esistenza di Dio*. Torino: Bollati Boringhieri.
- Gödel, K. (1999-2009). *Opere vol. 1-5*. Torino: Bollati Boringhieri.
- Goldman, A. (1967). A Causal Theory of Knowing. *The Journal of Philosophy* 64 , 357-372.
- Goldman, A. (1976). Discrimination and Perceptual Knowledge. *The Journal of Philosophy* 73 (20) , 771-791.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, Massachusetts: Harvard University Press.
- Gorgia. (2010). *Su Ciò che non è*. Hildesheim: Olms, Georg.
- Gould, S. J. (1997). Nonoverlapping Magisteria. *Natural History* 106 , 16-22, 60-62.
- Gould, S. J. (2002). *Rocks of Ages: Science and Religion in the Fullness of Life*. New York City: Ballantine.
- Gramsci, A. (2014). *Quaderni dal Carcere*. Torino: Einaudi.

- Greene, J. A., & Brown, S. C. (2006). The Wisdom Development Scale: Translating the Conceptual to the Concrete. *Journal of College Student Development*, 47, 1-19.
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. New York City: Penguin Press.
- Greene, J., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment. *Cognition*, 111 (3), 364-371.
- Grover, D., Camp, J., & Belnap, N. (1975). A Prosentential Theory of Truth. *Philosophical Studies* (27), 73-125.
- Grozio, U. (2002). *Il Diritto della Guerra e della Pace*. Firenze: Centro Editoriale Toscano.
- Haack, S. (1993). *Evidence and Inquiry*. Oxford: Blackwell.
- Habermas, J. (1993). *Etica del Discorso*. Roma-Bari: Laterza.
- Habermas, J. (2013). *Fatti e Norme. Contributi ad una Teoria Discorsiva del Diritto e della Democrazia*. Roma-Bari: Laterza.
- Habermas, J. (2001). *Verità e Giustificazione*. Roma-Bari: Laterza.
- Haraway, D. (2018). *Manifesto Cyborg. Donne, tecnologie e biopolitiche del corpo*. Milano: Feltrinelli.
- Hare, R. M. (1993). *Essays in Ethical Theory*. Oxford: Clarendon Press.
- Hare, R. M. (1963). *Freedom and Reason*. Oxford: Clarendon Press.
- Hare, R. M. (1981). *Moral Thinking: Its Levels, Method, and Point*. Oxford: Oxford University Press.
- Hare, R. M. (1952). *The Language of Morals*. Oxford: Clarendon Press.
- Harriet, H. (2011). *God, Goodness and Philosophy*. Farnham: Ashgate Publishing.
- Harris, S. (2012). *Il Paesaggio Morale*. Torino: Einaudi.
- Hauerwas, S. (2020). *Il Cristo Straziato. Le Ultime Parole di Cristo in Croce*. Brescia: Queriniana.
- Hauerwas, S. (1981). *Vision and Virtue: Essays in Christian Ethical Reflection*. Notre Dame, Indiana: University of Notre Dame Press.
- Hegel, G. W. (1807). *Fenomenologia dello Spirito*.
- Heidegger, M. (1927). *Essere e Tempo*.
- Henry, M. (1997). *Io Sono la Verità*. Brescia: Queriniana.
- Hintikka, J. (1975). *Models for Modalities: Selected Essays*. New York City: Springer.
- Hitchens, C. (2007). *Dio non è Grande. Come la Religione Avvelena ogni Cosa*. Torino: Einaudi.

- Hobbes, T. (1651). *Leviatano*.
- Hofstadter, D. R., & Dennett, D. (1991). *L'io della Mente*. Milano: Adelphi.
- Hölderlin, F. (2004). *Scritti di Estetica*. Milano: SE.
- Holt, E. B., Marvin, W. T., Montague, W. P., Perry, R. B., Pitkin, W. B., & Spaulding, E. G. (1912). *The New Realism: Cooperative Studies in Philosophy*. New York: The Macmillan Company.
- Hooker, B. (2000). *Ideal Code, Real World*. Oxford: Oxford University Press.
- Horgan, T., & Timmons, M. (1991). New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research*, 16 , 447-465.
- Horkheimer, M. (2000). *Eclisse della Ragione. Critica della Ragione Strumentale*. Torino: Einaudi.
- Horkheimer, M. (2003). *Filosofia e Teoria Critica*. Torino: Einaudi.
- Horton, J. (2017). The All or Nothing Problem. *Journal of Philosophy*, 114 , 94-104.
- Horwich, P. (2004). *From a Deflationary Point of View*. Oxford: Oxford University Press.
- Horwich, P. (1993). Gibbard's Theory of Norms. *Philosophy and Public Affairs*, 22 , 67-79.
- Horwich, P. (2010). *Truth -- Meaning -- Reality*. Oxford: Oxford University Press.
- Horwich, P. (1998). *Truth*. Oxford: Clarendon Press.
- Howard, M. C., & King, J. E. (1992). *A History of Marxian Economics: Volume II*. Princeton, New York: Princeton University Press.
- Huemer, M. (2007). Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research* 74 , 30-55.
- Huemer, M. (2011). Does Probability Theory Refute Coherentism? *Journal of Philosophy* 108 , 35-54.
- Huemer, M. (2005). *Ethical Intuitionism*. Londra: Palgrave MacMillan.
- Huemer, M. (1997). Probability and Coherence Justification. *Southern Journal of Philosophy* 35 , 463-472.
- Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, Maryland: Rowman & Littlefield.
- Huemer, M. (2007). Weak Bayesian Coherentism. *Synthese* 157 (3) , 337-346.
- Hughes, G. E., & Cresswell, M. G. (1996). *A New Introduction to Modal Logic*. Abingdon: Routledge.
- Hume, D. (2009). *Ricerca sui Principi della Morale*. Roma-Bari: Laterza.
- Hume, D. (1748). *Ricerca sull'Intelletto Umano*.

Hume, D. (1739). *Trattato sulla Natura Umana*.

Hursthouse, R., & Pettigrove, G. (2018). Virtue Ethics. In E. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2018 Edition)* (p. <https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/>).

Husserl, E. (1900). *Ricerche Logiche*.

Ippocrate. (1923). *Ancient Medicine. Airs, Waters, Places. Epidemics 1 and 3. The Oath. Precepts. Nutriment*. Cambridge, Massachusetts: Harvard University Press.

Irigaray, L. (1978). *Questo Sesso che non è un Sesso*. Milano: Feltrinelli.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Jackson, F., Oppy, G., & Smith, M. (1994). Minimalism and Truth Aptness. *Mind* 104 (411) , 287-302.

JafariNaimi, N. (2018). Our Bodies in the Trolley's Path, or Why Self-driving Cars Must *Not* Be Programmed to Kill. *Science, Technology, & Human Values*, 43 (2) , 302-323.

Jakobson, R. (1986). *La Scienza del Linguaggio*. Roma: Theoria.

James, W. (1984). *La Volontà di Credere*. Milano: Rizzoli.

James, W. (2008). *Pragmatismo*. Torino: Aragno.

James, W. (2009). *Saggi di Empirismo Radicale*. Macerata: Quodlibet.

Jameson, F. (2007). *Postmodernismo. Ovvero la Logica Culturale del Tardo Capitalismo*. Roma: Fazi.

Jarman, D. (Regia). (1993). *Wittgenstein* [Film].

Jaspers, K. (2006). *La Filosofia dell'Esistenza*. Roma-Bari: Laterza.

Johnson, R., & Cureton, A. (2021). Kant's Moral Philosophy. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Spring 2021 Edition)* (p. <https://plato.stanford.edu/archives/spr2021/entries/kant-moral/>).

Johnston, I. (. (2010). *The Mozi: a Complete Translation*. New York City: Columbia University Press.

Joyce, R. (2016). Moral Anti-Realism. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)* (p. <https://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/>).

Judea Pearl, A. P. (1986). Graphoids: A Graph-Based Logic for Reasoning About Relevance Relations. *ECAI'86: Proceedings of the 7th European Conference on Artificial Intelligence* , (p. Vol. 2: 357-363).

Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kahneman, D. (2010, Febbraio). The Riddle of Experience vs. Memory. *TED: Ideas Worth Spreading* . https://www.ted.com/talks/daniel_kahneman_the_riddle_of_experience_vs_memory.

- Kamm, F. (1996). *Morality, Mortality*. New York City: Oxford University Press.
- Kant, I. (1790). *Critica del Giudizio*.
- Kant, I. (1788). *Critica della Ragion Pratica*.
- Kant, I. (1781). *Critica della Ragion Pura*.
- Kant, I. (1785). *Fondazione della Metafisica dei Costumi*.
- Kant, I. (1797). *La Metafisica dei Costumi*.
- Kant, I. (1792). *La Religione Entro i Limiti della Semplice Ragione*.
- Kant, I. (2004). *Lezioni di Etica*. Roma-Bari: Laterza.
- Kelly, G. (2004). *La Psicologia dei Costrutti Personali. Teoria e Personalità*. Milano: Cortina Raffaello.
- Keplero, G. (1619). *Harmonices Mundi*.
- Keplero, G. (1596). *Mysterium Cosmographicum*.
- Khlentzos, D. Challenges to Metaphysical Realism. In E. N. (ed.), *Stanford Encyclopedia of Philosophy (Spring 2021 Edition)* (p. <https://plato.stanford.edu/archives/spr2021/entries/realism-sem-challenge/>).
- Kierkegaard, S. (1843). *Diario del Seduttore (in "Aut Aut")*.
- Kim, J. (. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Klein, P. D. (1998). Foundationalism and the Infinite Regress of Reasons. *Philosophy and Phenomenological Research* 58 , 919-925.
- Kleingeld, P. (2017). Contradiction and Kant's Formula of Universal Law. *Kant-Studien* 108 (1) , 89-115.
- Klima, G. (2017). The Medieval Problem of Universals. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)* (p. <https://plato.stanford.edu/archives/win2017/entries/universals-medieval/>).
- Korsgaard, C. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Kripke, S. (1999). *Nome e Necessità*. Torino: Bollati Boringhieri.
- Kripke, S. (2011). *Philosophical Troubles: Collected Papers*. Oxford: Oxford University Press.
- Kripke, S. (2000). *Wittgenstein su Regole e Linguaggio Privato*. Torino: Bollati Boringhieri.
- Kuhn, T. (2009). *La Struttura delle Rivoluzioni Scientifiche*. Torino: Einaudi.
- Lacan, J. (2002). *Scritti*. Torino: Einaudi.
- Laing, R. D. (1971). *La Politica dell'Esperienza*. Milano: Feltrinelli.

- Lakoff, G. (1999). *Donne, Fuoco e Cose Pericolose*. Firenze: La Nuova Italia.
- Lao-Tzu. (2013). *Il Libro del Tao. Tao-Teh-Ching*. Roma: Newton Compton.
- Latus, A. (2000). Moral and Epistemic Luck. *Journal of Philosophical Research*, 25 , 149-172.
- Lehrer, K. (1974). *Knowledge*. Oxford: Oxford University Press.
- Lehrer, K. (1990). *Theory of Knowledge*. Abingdon-on-Thames: Routledge.
- Lehrer, K., & Paxson, T. (1969). Knowledge: Undefeated Justifies True Belief. *The Journal of Philosophy* 66 (8) , 225-237.
- Leibniz, G. (1666). *Disputatio Inauguralis de Casibus Perplexis in Jure*.
- Leibniz, G. (1720). *Monadologia*.
- Leibniz, G. (1765). *Nuovi Saggi sull'Intelletto Umano*.
- Leibniz, G. (1710). *Teodicea*.
- Lemmon, E. J. (1986). *Elementi di Logica*. Roma-Bari: Laterza.
- Levi Bryant, N. S. (2011). *The Speculative Turn: Continental Materialism and Realism*. Melbourne: re.press.
- Lévinas, E. (2016). *Totalità e Infinito. Saggio sull'Esteriorità*. Milano: Jaca Book.
- Lévinas, E., Marcel, G., & Ricoeur, P. (2008). *Il Pensiero dell'Altro*. Roma: Edizioni Lavoro.
- Lévi-Strauss, C. (2008). *Tristi Tropici*. Milano: Il Saggiatore.
- Lewis, C. I. (1946). *An Analysis of Knowledge and Valutation*. La Salle, Illinois: Open Court.
- Lewis, C. I., & Langford, C. H. (1932). *Symbolic Logic*. New York City: Century Company.
- Lewis, D. K. (1973). Causation. *Journal of Philosophy*, 70 (17) , 556-567.
- Lewis, D. K. (2000). Causation as Influence. *Journal of Philosophy*, 97 (4) , 182-197.
- Lewis, D. K. (1970). *How to Define Theoretical Terms*. *Journal of Philosophy*, 67: 426-446.
- Lewis, D. K. (1984). Putnam's Paradox. *Australasian Journal of Philosophy* 62 , 221-236.
- Lewis, D. K. (2020). *Sulla Pluralità dei Mondi*. Milano: Mimesis.
- Lewis, D. K. (1976). The Paradoxes of Time Travel. *American Philosophical Quarterly*, 13 , 145-152.
- Locke, J. (1690). *Saggio sull'Intelletto Umano*.
- Lockie, R. (2018). *Free Will and Epistemology: A Defence of the Transcendental Argument for Freedom*. Bloomsbury Academic.
- Logic, F. o. (1985). *John Searle; Daniel Vanderveken*. Cambridge: Cambridge University Press.

Lokhorst, G.-J. (2019). Mally's Deontic Logic. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)* (p. <https://plato.stanford.edu/archives/sum2019/entries/mally-deontic/>).

Loneragan, B. (2001). *Il Metodo in Teologia*. Roma: Città Nuova.

Lucas, R. E., & Lawless, N. M. (2013). Does Life Seem Better on a Sunny Day? Examining the Association Between Daily Weather Conditions and Life Satisfaction Judgments. *Journal of Personality and Social Psychology*, 104 (5) , 872-884.

Lukács, G. (1997). *Storia e Coscienza di Classe*. Milano: Sugarco.

Lukács, G. (2004). *Teoria del Romanzo*. Milano: SE.

Machiavelli, N. (1532). *Il Principe*.

MacIntyre, A. (2007). *Dopo la Virtù. Saggio di Teoria Morale*. Roma: Armando.

Mackie, J. L. (2001). *Etica: Inventare il Giusto e l'Ingiusto*. Torino: Giappichelli.

Maimonide, M. (2013). *La Guida dei Perplexi*. Milano: UTET.

Mangan, J. (1949). An Historical Analysis of the Principle of Double Effect. *Theological Studies*, 10 , 41-61.

Marco-Aurelio. (2010). *Ricordi*. Milano: Rizzoli.

Marconi, D. (2008). *La Filosofia del Linguaggio. Da Frege ai Giorni Nostri*. Torino: UTET.

Marcuse, H. (2001). *Eros e Civiltà*. Torino: Einaudi.

Marion, J.-L. (2018). *Dio Senza Essere*. Milano: Jaca Book.

Marx, K. (1867). *Il Capitale*.

Marx, K. (2018). *Scritti Economico-Filosofici del 1844. E Altre Pagine su Lavoro e Alienazione*. Milano: Feltrinelli.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional Intelligence: Theory, Findings, and Implications. *Psychological Inquiry*, 60 (3) , 197-215.

McConnell, T. (1989). "Ought" Implies "Can" and the Scope of Moral Requirements. *Philosophia*, 19 , 437-454.

McDowell, J. (1998). *Mind, Value, and Reality*. Oxford: Oxford University Press.

McGrath, A. (2016). *La Grande Domanda*. Torino: Bollati Boringhieri.

Meinong, A. (2019). *Teoria dell'Oggetto*. Macerata: Quodlibet.

Merleau-Ponty, M. (2003). *Fenomenologia della Percezione*. Milano: Bompiani.

Miche, G. (1946). *Manuale di Filosofia Bolscevica*. Roma: Magi-Spinetti.

Mill, J. S. (1976). *Autobiografia*. Roma-Bari: Laterza.

- Mill, J. S. (1840). *Bentham e Coleridge. Due Saggi*.
- Mill, J. S. (1991). *L'Utilitarismo*. Milano: SugarCo.
- Mills, E. (1998). A Simple Solution to the Liars. *Philosophical Studies* 89 (2/3) , 197-212.
- Misak, C. (2020). *Frank Ramsey: A Sheer Excess of Powers*. Oxford: Oxford University Press.
- Mitchell, D. (1971). *Storia della Sociologia Moderna*. Milano: Mondadori.
- Moderno, I. M. (1997). *Isaiah Berlin*. Milano: Adelphi.
- Monk, R. (2000). *Wittgenstein. Il Dovere del Genio*. Milano: Bompiani.
- Montesquieu. (1748). *Lo Spirito delle Leggi*.
- Moore, G. E. (1925). *A Defence of Common Sense*.
- Moore, G. E. (1993). Moore's Paradox. In T. (. Baldwin, G. E. Moore: *Selected Writings* (p. 207-212). London: Routledge.
- Moore, G. E. (1903). *Principia Ethica*.
- More, T. (2016). *Utopia*. Milano: Feltrinelli.
- Morris, M. (2005). Realism Beyond Correspondence. In H. Beebee, & J. Dodd, *Truthmakers: The Contemporary Debate* (p. ch. 4). Oxford: Oxford University Press.
- Moser, P. (1985). *Empirical Justification*. Dordrecht: Springer.
- Moser, P. (1991). *Knowledge and Evidence*. Cambridge: Cambridge University Press.
- Mounier, E. (2004). *Il Personalismo*. Roma: AVE.
- Nagel, T. (2015). *Mente e Cosmo. Perché la Concezione Neodarwiniana della Natura è Quasi Certamente Falsa*. Milano: Raffaello Cortina.
- Nagel, T. (1991). *Moral Questions*. Cambridge: Cambridge University Press.
- Nakano, A. (2020). Anti-Realism and Anti-Revisionism in Wittgenstein's Philosophy of Mathematics. *Grazer Philosophische Studien* 97 (3) , 451-474.
- Nedo, M. (. (2018). *Wittgenstein. Una Biografia per Immagini*. Roma: Carocci.
- Nicolai, C., & Cohnitz, D. (inedito). *How to Adopt a Logic*. Tratto il giorno febbraio 18, 2021 da https://carlonicolai.github.io/Adop9_website.pdf
- Nietzsche, F. (1886). *Al di là del Bene e del Male*.
- Nietzsche, F. (1883). *Così Parlò Zarathustra*.
- Nozick, R. (2005). *Anarchia, Stato e Utopia*. Milano: Il Saggiatore.
- Nozick, R. (2004). *La Vita Pensata. Meditazioni Filosofiche*. Milano: Rizzoli.

- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, Massachusetts: Harvard University Press.
- Nozick, R. (1999). *Puzzle Socratici*. Milano: Raffaello Cortina.
- Nussbaum, M. C. (2000). *Women and Human Development: The Capabilities Approach*. New York City: Cambridge University Press.
- Nussbaum, M. (2011). *La Fragilità del Bene. Fortuna ed Etica nella Tragedia e nella Filosofia Greca*. Bologna: Il Mulino.
- Nussbaum, M. (1997). *Terapia del Desiderio. Teoria e Pratica nell'Etica Ellenistica*. Milano: Vita e Pensiero.
- Nussbaum, M. (1999). Virtue Ethics: A Misleading Category? *The Journal of Ethics*, 3 (3) , 163-201.
- Nussbaum, M., & Sen, A. K. (1993). *The Quality of Life*. Oxford: Clarendon Press.
- Odifreddi, P. (2011). *Caro Papa, ti Scrivo*. Milano: Mondadori.
- Odifreddi, P. (2018). *Il Dio della Logica. Vita Geniale di Kurt Gödel, Matematico della Filosofia*. Milano: Longanesi.
- Olsson, E. (2017). Coherentist Theories of Epistemic Justification. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*. <https://plato.stanford.edu/archives/spr2017/entries/justep-coherence/>.
- Olsson, E. (2002). What is the Problem of Coherence and Truth? *The Journal of Philosophy* 99 , 246-272.
- Onfray, M. (2006-2018). *Contre-histoire de la Philosophie, vol. 11*. Paris: Grasset.
- Onfray, M. (2007). *La Scultura di Sé. Per una Morale Estetica*. Roma: Fazi.
- Origene. (2012). *Commento al Vangelo di Giovanni*. Milano: Bompiani.
- Paci, E. (. (1971). *Bertrand Russell*. Milano: Club degli Editori.
- Palmira, M. (2014). Is Rational Disagreement in Philosophy Possible? *Iride*, 3 , 595-614.
- Panezio. (1997). *Testimonianze*. Napoli: Bibliopolis.
- Papineau, D. (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Papini, G. (2018). *Il Crepuscolo dei Filosofi. Kant, Hegel, Schopenhauer, Comte, Spencer, Nietzsche*. Milano: GOG.
- Parfit, D. (1982). Future Generations: Further Problems. *Philosophy and Public Affairs*, 11 , 113-172.
- Parfit, D. (1989). *Ragioni e Persone*. Roma: Theoria.
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

- Pedersen, N. (2010). Stabilizing Alethic Pluralism. *Philosophical Quarterly* (60) , 92-108.
- Pedersen, N. (2006). What can the problem of mixed inferences teach us about alethic pluralism? *The Monist* (89) , 103-117.
- Peirce, C. S. (2011). *Opere*. Milano: Bompiani.
- Perry, K. (2015). *Philosophy: an Illuminating Guide to History's Greatest Thinkers*. New York City: Sterling.
- Peterson, J. (2018). *12 Regole per la Vita. Un Antidoto al Caos*. Coriano: My Life.
- Piaget, J. (2016). *Logica e Conoscenza Scientifica*. Roma: Studium.
- Pietro-Mantovano. (1492). *Logica*. Venezia: Boneto Locatello.
- Pigden, C. (2016). Hume On Is and Ought: Logic, Promises and the Duke of Wellington. In P. Russell, *The Oxford Handbook of David Hume*. New York City: Oxford University Press.
- Pigliucci, M. (2017). *Come Essere Stoici. Riscoprire la Spirituaità degli Antichi per Vivere una Vita Moderna*. Milano: Garzanti.
- Pinker, S. (1998). *L'Istinto del Linguaggio*. Milano: Mondadori.
- Pirrone. (1981). *Testimonianze*. Napoli: Bibliopolis.
- Plantinga, A. (2001). *Essays in the Metaphysics of Modality*. Oxford: Oxford University Press.
- Plantinga, A. (1967). *God and Other Minds*. Ithaca, New York: Cornell University Press.
- Plantinga, A. (1974). *The Nature of Necessity*. Londra: Oxford University Press.
- Plantinga, A. (1991). When Faith and Reason Clash: Evolution and the Bible. *Christian Scholars Review* 21 (1) , 8-32.
- Platone. (2009). *Tutte le Opere*. Roma: Newton Compton.
- Pnueli, A. (1977). The Temporal Logic of Programs. *Proceedings of the 6th Symposium on the Foundations of Computer Science* , 46-57.
- Polanyi, M. (2012). *Fede e Ragione*. Brescia: Morcelliana.
- Polanyi, M. (2018). *La Conoscenza Inespressa*. Roma: Armando.
- Polkinghorne, J. (2000). *Credere in Dio nell'Età della Scienza*. Milano: Raffaello Cortina.
- Pomponazzi, P. (2013). *Tutti i Trattati Peripatetici*. Milano: Bompiani.
- Popper, K. (2009). *Congetture e Confutazioni. Lo Sviluppo della Conoscenza Scientifica*. Bologna: Il Mulino.
- Popper, K. (2010). *Logica della Scoperta Scientifica*. Torino: Einaudi.
- Posidonio. (2004). *Testimonianze e Frammenti*. Milano: Bompiani.

- Poston, T. (2014). *Reason and Explanation: A Defense of Explanatory Coherentism*. New York City: Palgrave Macmillan.
- Priest, G. (1998). What is so Bad about Contradictions? *Journal of Philosophy* 95 , 410-426.
- Prior, A. N. (1976). *Papers in Logic and Ethics*. London: Duckworth.
- Prior, A. N. (2003). *Time and Modality*. Oxford: Clarendon Press.
- Proclo. (1994). *Elementi di Teologia*. Firenze: La Nuova Italia.
- Putnam, H. (2004). *Fatto-Valore. Fine di una Dicotomia*. Roma: Fazi.
- Putnam, H. (1993). *Matematica, Materia, Metodo*. Milano: Adelphi.
- Putnam, H. (1987). *Mente, Linguaggio e Realtà*. Milano: Adelphi.
- Putnam, H. (1980). Models and Reality. *Journal of Symbolic Logic* 45 (3) , 464-482.
- Putnam, H. (1985). *Ragione, Verità e Storia*. Milano: Il Saggiatore.
- Putnam, H. (1977). Realism and Reason. *Proceedings and Addresses of the American Philosophical Association* 50 (6) , 483-498.
- Quine, W. v. (2004). *Da un Punto di Vista Logico. Saggi Logico-Filosofici*. Milano: Raffaello Cortina.
- Quine, W. v. (2008). *Parola e Oggetto*. Milano: Il Saggiatore.
- Quine, W. V. (1986). *Philosophy of Logic*. Cambridge, Massachusetts: Harvard University Press.
- Quine, W. v. (1951). Two Dogmas of Empiricism. *The Philophical Review* , 60 (1): 20-43.
- Quinn, W. (1989). Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing. *Philosophical Review*, 98 (3) , 287-312.
- Rachels, J. (1975). Active and Passive Euthanasia. *New England Journal of Medicine*, 292 , 78-86.
- Rahner, K. (1992). *La Penitenza della Chiesa. Saggi Teologici e Storici*. Cinisello Balsamo: San Paolo.
- Ramsey, F. P. (1964). *I Fondamenti della Matematica e Altri Scritti di Logica*. Milano: Feltrinelli.
- Ramsey, F. P. (2020). *Sulla Verità e Scritti Pragmatisti*. Macerata: Quodlibet.
- Rand, A. (2004). *La Fonte Meravigliosa*. Milano: Corbaccio.
- Rand, A. (1999). *La Virtù dell'Egoismo*. Macerata: Liberilibri.
- Rawls, J. (2008). *Una Teoria della Giustizia*. Milano: Feltrinelli.
- Reich, S. (Compositore). (2006). *Daniel Variations*.
- Reid, T. (1996). *Ricerca sulla Mente Umana e Altri Scritti*. Torino: UTET.

- Rescher, N. (2005). *Epistemic Logic: A Survey Of the Logic Of Knowledge*. Pittsburgh, Pennsylvania: University of Pittsburgh Press.
- Rescher, N. (1979). Russell and Modal Logic. In G. Roberst, *Bertrand Russell Memorial Volume*. Londra: Allen & Unwin.
- Rescher, N. (1982). *The Coherence Theory of Truth*. Washington, D.C.: University Press of America.
- Ricard, M. (2015). *Happiness: A Guide to Developing Life's Most Important Skill*. Londra: Atlantic Books.
- Richards, N. (1986). Moral Luck and Desert. *Mind*, 65 , 198-209.
- Ricoeur, P. (2016). *Dall'Azione al Testo. Saggi di Ermeneutica*. Milano: Jaca Book.
- Roberts, M. A. (1998). *Child Versus Childmaker: Future Persons and Present Duties in Ethics and The Law*. Lanham, Maryland: Rowman & Littlefield.
- Rogers, B. (1999). *A. J. Ayer: A Life*. New York City: Grove Press.
- Rorty, R. (2004). *La Filosofia e lo Specchio della Natura*. Milano: Bompiani.
- Ross, W. D. (1939). *Foundations of Ethics: The Gifford Lectures Delivered in the University of Aberdeen*. Oxford: Oxford University Press.
- Ross, W. D. (1930). *The Right and the Good. Some Problems in Ethics*. Oxford: Clarendon Press.
- Russell, B. (2009). *Human Knowledge: its Scope and its Limits*. Abingdon-on-Thames: Routledge.
- Russell, B. (2013). *I Problemi della Filosofia*. Milano: Feltrinelli.
- Russell, B. (2003). *La Filosofia dell'Atomismo Logico*. Torino: Einaudi.
- Russell, B. (2012). *La Saggezza dell'Occidente*. Milano: TEA.
- Russell, B. (1905). On Denoting. *Mind (14)* , 479-493.
- Russell, B. (1907). On the Nature of Truth. *Proceedings of the Aristotelian Society (7)* , 228-249.
- Russell, B. (1994). *Philosophical Essays*. London - New York: Routledge.
- Russell, B. (2004). *Storia della Filosofia Occidentale dei suoi Rapporti con le Vicende Politiche e Sociali dall'Antichità a Oggi*. Milano: TEA.
- Rutten, E. (2014, agosto 28). A semantic argument against the existence of universally held real properties. (*conferenza*) . Vrije Universiteit Amsterdam.
- Ryberg, J., & (ed.), T. T. (2004). *The Repugnant Conclusion. Essays on Population Ethics*. Dordrecht: Kluwer Academic Publishers.
- Salerno, J. (2009). *New Essays on the Knowability Paradox*. Oxford: Oxford University Press.

- Salomon, D. (1988). Internal Objections to Virtue Ethics. *Midwest Studies in Philosophy*, 13 , 428-441.
- Sartre, J.-P. (1938). *La Nausea*.
- Sartre, J.-P. (1946). *L'Esistenzialismo è un Umanismo*.
- Savavarsdottir, S. (1999). Moral Cognitivism and Motivation. *Philosophical Review*, 108 , 161-219.
- Sayre-McCord, G. (1997). "Good" on Twin Earth. *Philosophical Issues*, 8 , 267-292.
- Scheler, M. (2013). *Il Formalismo nell'Etica e l'Etica Materiale dei Valori*. Milano: Bompiani.
- Schelling, F. (1990). *Scritti sulla Filosofia, la Religione, la Libertà*. Milano: Mursia.
- Schelling, F. W. (1801). *Esposizione del mio Sistema Filosofico*.
- Schelling, F. W. (1988). *Ideas for a Philosophy of Nature: as Introduction to the Study of this Science*. Cambridge: Cambridge University Press.
- Schleiermacher, F. D. (2000). *Ermeneutica*. Milano: Bompiani.
- Schleiermacher, F. D. (1985). *Etica ed Ermeneutica*. Napoli: Bibliopolis.
- Schlick, M. (1936). Meaning and Verification. *Philosophical Review* 45 (4) , 339-369.
- Schopenhauer, A. (2011). *Il Mondo come Volontà e Rappresentazione*. Roma: Newton Compton.
- Schroeder, M. (2008). How Expressivists Can and Should Solve Their Problem With Negation. *Noûs*, 42 (4) , 573-599.
- Schulze, G. E. (1792). *Aenesidemus*.
- Schütz, A. (2018). *La Fenomenologia del Mondo Sociale*. Roma: Meltemi.
- Schwitzgebel, E. (2019). The Stanford Encyclopedia of Philosophy (Fall 2019 Edition). In E. N. Zalta.
- Scoto, G. D. (2005). *Ordinatio. Commento alle Sentenze*. Frigento: Casa Mariana.
- Scoto, G. D. (2008). *Trattato sul Primo Principio*. Milano: Bompiani.
- Scruton, R. (2015). *Essere Conservatore*. Crotone: D'Ettoris.
- Scruton, R. (2007). *Manifesto dei Conservatori*. Milano: Raffaello Cortina.
- Scruton, R. (2018). *Sulla Natura Umana*. Milano: Vita e Pensiero.
- Searle, J. (2009). *Atti Linguistici. Saggi di Filosofia del Linguaggio*. Torino: Bollati Boringhieri.
- Searle, J. (2010). *Creare il Mondo Sociale. La Struttura della Civiltà Umana*. Milano: Raffaello Cortina.

Searle, J. (1985). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Searle, J. (1964). How to Derive "Ought" from "Is". *The Philosophical Review* 73 (1) , 43-58.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, J. (2006). *La Costruzione della Realtà Sociale*. Torino: Einaudi.

Searle, J. (2003). *La Razionalità dell'Azione*. Milano: Raffaello Cortina.

Segal, L. (2001). *The Dream of Reality*. New York City: Springer.

Seligman, M. (2017). *Fai Fiorire la tua Vita. Una Nuova, Rivoluzionaria Visione della Felicità e del Benessere*. Torino: Anteprema.

Seligman, M. (2007). *La Costruzione della Felicità*. Milano: Fabbri.

Seligman, M., & Peterson, C. (2004). *Character Strengths and Virtues*. Oxford: Oxford University Press.

Sellars, W. (2004). *Empirismo e Filosofia della Mente*. Torino: Einaudi.

Sellars, W. (2007). *La Filosofia e l'Immagine Scientifica dell'Uomo*. Roma: Armando Editore.

Sellars, W. (2013). *L'Immagine Scientifica e l'Immagine Manifesta*. Pisa: ETS.

Sen, A. (1985). Well-Being, Agency, and Freedom. *Journal of Philosophy*, 82 (4) , 169-221.

Seneca. (2015). *Tutte le Opere*. Roma: Newton Compton.

Sesto-Empirico. (1999). *Schizzi Pirroniani*. Milano: Lampi di Stampa.

Šestov, L. (2005). *Atene e Gerusalemme*. Milano: Bompiani.

Šestov, L. (2016). *Parmenide Incatenato*. Milano: Luni Editrice.

Shankaracharya. (2015). *The Crest-Jewel of Wisdom*. Wolcott, New York: Scholar's Choice.

Sidgwick, H. (1981). *The Methods of Ethics*. Indianapolis: Hackett.

Sikora, R. I., & Barry, B. (1978). *Obligations to Future Generations*. Philadelphia: Temple University Press.

Simon, H. A. (1988). *Le Scienze dell'Artificiale*. Milano: Feltrinelli.

Sinclair, T. (2018). Are We Conditionally Obligated to Be Effective Altruists? *Philosophy and Public Affairs*, 46 , 36-59.

Singer, P. (1995). *Etica Pratica*. Napoli: Liguori.

Singer, P. (2016). *La Cosa Migliore che Puoi Fare. Cos'è l'Altruismo Efficace*. Milano: Sonda.

Singer, P., & de Lazari-Radek, K. (2014). *The Point of View of the Universe*. Oxford: Oxford University Press.

Skolem, T. (1977). Some Remarks on Axiomatized Set Theory. In J. v. (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931* (p. 290-301). Cambridge, Massachusetts: Harvard University Press.

Slote, M. (1984). Satisficing Consequentialism. *Proceedings of the Aristotelian Society*, 58 , 139-163.

Smeenk, C., & Ellis, G. (2017). Philosophy of Cosmology. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)* (p. <https://plato.stanford.edu/archives/win2017/entries/cosmology/>).

Smith, B. (1982). *Parts and Moments. Studies in Logic and Formal Ontology*. Munich: Philosophia.

Smith, M. (1987). The Humean Theory of Motivation. *Mind* 96 (381) , 36-61.

Smullyan, R. (1987). *Forever Undecided*. New York City: Alfred A. Knopf.

Sokal, A. (1999). *Imposture Intellettuali. Quale deve Essere il Rapporto tra Filosofia e Scienza?* Milano: Feltrinelli.

Sokal, A. (1996). Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity. *Social Text* 46-47 , 217-252.

Sosa, E. (1999). How to Defeat Opposition to Moore. *Noûs* 33 (s13) , 141-153.

Spaventa, B. (2009). *Opere*. Milano: Bompiani.

Spinoza, B. (1677). *L'Etica*.

Sternberg, R. J. (2003). *Wisdom, Intelligence, and Creativity Synthesized*. New York City: Cambridge University Press.

Sternberg, R. J. (1990). *Wisdom: Its Nature, Origins, and Development*. New York City: Cambridge University Press.

Stevens, M. (2017, dicembre 6). The Greater Good - Mind Field S2 (Ep 1). <https://www.youtube.com/watch?v=1sl5KJ69qiA>.

Stevenson, C. L. (1962). *Etica e Linguaggio*. Milano: Longanesi.

Stigler, S. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, Massachusetts: Harvard University Press.

Stirner, M. (1999). *L'Unico e la Sua Proprietà*. Milano: Adelphi.

Stocker, M. (1976). The Schizophrenia of Modern Ethical Theories. *The Journal of Philosophy*, 73 (14) , 453-466.

Stoljar, D., & Damjanovic, N. (2014). The Deflationary Theory of Truth. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Fall 2014 Edition)*. <https://plato.stanford.edu/archives/fall2014/entries/truth-deflationary/>.

Stoutland, F. (2003). What philosophers should know about truth and the slingshot. In M. Sintonen, P. Ylikoski, & K. M. (ed.), *Realism in Action: Essays in the Philosophy of the Social Sciences* (p. 3-32). Dordrecht: Kluwer Academic Publishers.

Stratton-Lake, P. (2020). Intuitionism in Ethics. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)* (p. <https://plato.stanford.edu/archives/sum2020/entries/intuitionism-ethics/>).

Strawson, P. F. (1949). Truth. *Analysis* (9) , 83-97.

Sturgeon, N. (1988). Moral Explanations. In G. Sayre-McCord, *Essays on Moral Realism* (p. 229-255). Ithaca-London: Cornell University Press.

Swanton, C. (2003). *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press.

Szasz, T. (2003). *Il Mito della Malattia Mentale*. Bologna: Spirali.

Tappolet, C. (1997). Mixed inferences: a problem for pluralism about truth predicates. *Analysis* 57 , 209-210.

Tappolet, C. (2004). Through Thick and Thin: Good and its Determinates. *Dialectica*, 58 (2) , 207-221.

Tappolet, C. (2000). Truth Pluralism and Many-Valued Logic: a Reply to Beall. *Philosophical Quarterly* (50) , 382-385.

Tarski, A. (1983). The Concept of Truth in Formalized Language. In J. C. (ed.), *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (p. 152-278). Indianapolis: Hackett Publishing Company.

Taylor, B. (1985). *Modes of Occurrence: Verbs, Adverbs and Events*. Oxford: Blackwell.

Telesio, B. (2009). *La Natura Secondo i Suoi Principi*. Milano: Bompiani.

Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, Massachusetts: MIT Press.

Thagard, P. (2012, dicembre 4). *Eleven Dogmas of Analytic Philosophy*. Tratto il giorno febbraio 14, 2020 da Psychology Today: <https://www.psychologytoday.com/us/blog/hot-thought/201212/eleven-dogmas-analytic-philosophy>

Thagard, P. (2019). *Natural Philosophy: From Social Brains to Knowledge, Reality, Morality, and Beauty*. Oxford: Oxford University Press.

Thagard, P. (2005). Testimony, Credibility and Explanatory Coherence. *Erkenntnis* 63 (3) , 295-316.

Thomson, J. (1985). The Trolley Problem. *Yale Law Journal*, 94 (6) , 1395-1415.

Thomson, J. (2008). Turning the Trolley. *Philosophy and Public Affairs*, 36 (4) , 359-374.

Tillich, P. (1996-2006). *Teologia Sistemativa Vol. 1-4*. Torino: Claudiana.

Tolstoj, L. (2009). *Confessione*. San Pietro Capofiume: Ginko.

Toraldo di Francia, G. (1976). *L'Indagine del Mondo Fisico*. Torino: Einaudi.

- Turri, J., & Klein, P. (. (2014). *Ad Infinitum: New Essays on Epistemological Infinitism*. Oxford: Oxford University Press.
- Týn, T. (1986, novembre 13). Conferenza: L'Esistenza di Dio. Bologna: Istituto Tincani.
- Unger, P. (1992). Causing and Preventing Serious Hard. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 65 , 227-255.
- Unwin, N. (1999). Quasi-Realism, Negation and the Frege-Geach Problem. *Philosophical Quarterly*, 49 , 337-352.
- Vailati, G. (1980). *Scritti Filosofici*. Firenze: La Nuova Italia.
- Valerio, G. (2017). *Genesi ed Evoluzione della Matematica*. Tricase: Youcanprint Self-Publishing.
- Van Cleve, J. (2011). Can Coherence Generate Warrant Ex Nihilo? Probability and the Logic of Concurring Witnesses. *Philosophy and Phenomenological Research* 82(2) , 337-380.
- Varden, H. (2010). Kant and Lying to the Murderer at the Door... One More Time: Kant's Legal Philosophy and Lies to Murderers and Nazis. *Journal of Social Philosophy*, 41 (4) , 403-421.
- Väyrynen, P. (2013). *The Lewd, the Rude and the Nasty*. New York City: Oxford University Press.
- Väyrynen, P. Thick Ethical Concepts. In E. N. Zalta, *The Stanford Encyclopedia of Philosophy (Spring 2021 Edition)* (p. <https://plato.stanford.edu/archives/spr2021/entries/thick-ethical-concepts/>).
- Vico, G. (1725). *Scienza Nuova*.
- Vidor, K. (Regia). (1949). *La Fonte Meravigliosa* [Film].
- Viggiano, A. (2008). Ethical Naturalism and Moral Twin Earth. *Ethical Theory and Moral Practice*, 11 (2) , 213-224.
- Vitale, M. (2014). *Dalla Volontà di Vivere all'Inconscio. Eduard von Hartmann e la Trasformazione della Filosofia di Schopenhauer*. Lecce: Pensa Multimedia.
- von Glaserfeld, E. (2016). *Il Costruttivismo Radicale. Una Via per Conoscere e Apprendere*. Roma: Odradek.
- von Helmholtz, H. (2003). *Opere*. Torino: UTET.
- von Hildebrand, D. (2003). *Essenza dell'Amore*. Milano: Bompiani.
- Walker, R. C. (1989). *The Coherence Theory of Truth: Realism, Anti-Realism, Idealism*. London - New York: Routledge.
- Watzlawick, P. (. (2018). *La Realtà Inventata: Contributi al Costruttivismo*. Milano: Feltrinelli.
- Welles, O. (1938, ottobre 30). The War of the Worlds. New York City: CBS Radio.
- Wessels, U. (2015). Beyond the Call of Duty: The Structure of a Moral Region. *Royal Institute of Philosophy Supplement*, 77 , 87-104.
- Whitehead, A. N. (2015). *La Scienza e il Mondo Moderno*. Torino: Bollati Boringhieri.

- Whitehead, A. N. (1929). *Processo e Realtà*.
- Wiener, N., Rosenblueth, A., & Bigelow, J. (1943). Behavior, Purpose and Teleology. *Philosophy of Science* (10: 1) .
- Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, Massachusetts: Harvard University Press.
- Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, B. (1995). Truth in Ethics. *Ratio*, 8 (3) , 227-242.
- Williams, B., & Sen, A. K. (1982). *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
- Williamson, T. (1982). Intuitionism Disproved? *Analysis* (42) , 203-207.
- Wittgenstein, L. (2005). *Conversazioni e Ricordi*. Vicenza: Neri Pozza.
- Wittgenstein, L. (1999). *Della Certezza*. Torino: Einaudi.
- Wittgenstein, L. (1967). *Lezioni e Conversazioni sull'Etica, l'Estetica, la Psicologia e la Credenza Religiosa*. Milano: Adelphi.
- Wittgenstein, L. (2000). *Libro Blu e Libro Marrone*. Torino: Einaudi.
- Wittgenstein, L. (2014). *Ricerche Filosofiche*. Torino: Einaudi.
- Wittgenstein, L. (2009). *Tractatus Logico-Philosophicus e Quaderni 1914-1916*. Torino: Einaudi.
- Wittgenstein, L. (2021). *Vostro Fratello Ludwig*. Sesto San Giovanni: Mimesis.
- Wittgenstein, L. (1981). *Osservazioni sui Colori*. Torino: Einaudi.
- Wolff, C. (2011). *Metafisica Tedesca*. Milano: Bompiani.
- Wright, C. (1996). Response to Commentators. *Philosophy and Phenomenological Research* (56) , 911-941.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, Massachusetts: Harvard University Press.
- Wright, C. (2004). Warrant for Nothing (and Foundations for Free)? *Aristotelian Society Supplementary Volume 78* , 167-212.
- Zagzebski, L. (1994). The Inescapability of Gettier Problems. *The Philosophical Quarterly* 44 (174) , 65-73.
- Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.
- Zagzebski, L., & (ed.), M. D. (2003). *Intellectual Virtue: Perspectives from Ethics and Epistemology*. Oxford: Clarendon Press.
- Zagzebski, L., & Fairweather, A. (2001). *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. New York City: Oxford University Press.

Zalta, E. N. (1983). *Abstract Objects: an Introduction to Axiomatic Metaphysics*. New York City: Springer.

Zanetti, L. (2019). Determinism and Judgement: A Critique of the Indirect Epistemic Transcendental Argument for Freedom. *European Journal of Analytic Philosophy* , 15(2):33-54.

Žižek, S. (2016). *Il Contraccolpo Assoluto. Per una Nuova Fondazione del Materialismo Dialettico*. Milano: Ponte delle Grazie.

Žižek, S. (2012). *L'Isterico Sublime. Psicanalisi e Filosofia*. Sesto San Giovanni: Mimesis.

Žižek, S. (2017). *Meno di Niente. Hegel e l'Ombra del Materialismo Dialettivo Vol. 1-2*. Milano: Ponte delle Grazie.