LA MENTE E I SISTEMI COGNITIVI

Collana di scienze cognitive, filosofia e tecnologia

2

LA MENTE E I SISTEMI COGNITIVI

Collana di scienze cognitive, filosofia e tecnologia

Humani nihil a me alienum puto.
— Publio Terenzio Afro

La collana raccoglie e presenta testi scientifici che studiano i fenomeni mentali e sociali in differenti ambiti disciplinari (filosofia, psicologia, biologia, informatica, robotica, etica, linguistica, antropologia, ecc.). Ciò con l'obiettivo di mettere in luce le complesse relazioni che intercorrono fra cognizione, corpo, ambiente tecnologico e sociale, nonché le implicazioni etiche che derivano dallo sviluppo delle nuove tecnologie cognitive.

I limiti epistemologici degli studi disciplinari non consentono di elaborare una visione coerente sul funzionamento della mente. Di conseguenza, si pone la necessità di un quadro interdisciplinare più ampio, che favorisca l'interazione fra i vari ambiti disciplinari e l'integrazione delle varie prospettive di studio.

In questo senso, i testi della collana si devono intendere come contributi a un'impresa collettiva che cerca di colmare il divario fra le domande, sempre più incalzanti, che ci poniamo sulla natura e sul funzionamento della mente e le risposte parziali offerte dalle singole discipline.

Concetti e processi di categorizzazione

a cura di
Francesco Gagliardi
Marco Cruciani
Andrea Velardi

Contributi di
Angelo Cangelosi
Romolo Giovanni Capuano
Luciano Celi
Antonio Chella
Marco Cruciani
Edoardo Fugali
Francesco Gagliardi
Giuliana Gerace
Alberto Greco
Elisabetta Lalumera
Stefania Moretti
Alessio Plebe
Andrea Velardi

Aracne editrice

www.aracneeditrice.it
info@aracneeditrice.it

Gioacchino Onorati editore S.r.l. – unipersonale

www.gioacchinoonoratieditore.it
info@gioacchinoonoratieditore.it

via Vittorio Veneto, 20
00020 Canterano (RM)
(06) 45551463

ISBN 978-88-255-1306-6

I edizione: marzo 2018

Indice

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130661
pp. 9-13 (marzo 2018)

Prefazione

di Francesco Gagliardi[1],
Marco Cruciani[2], Andrea Velardi[3]

In questo volume sono raccolti undici saggi che discutono alcuni dei temi più innovativi nell'area di ricerca relativa alle teorie dei concetti e dei processi di categorizzazione anche in relazione con altre funzioni e processi cognitivi e tenendo intrecciate la dimensione più generale, teorica con quella, più specifica, di approfondimento tematico e di natura applicativa. Il volume si presenta così come una raccolta di carattere multidisciplinare e multisettoriale che analizza scenari consolidati, contesti significativi già noti alla letteratura e al dibattito internazionali e che non esita però ad avventurarsi nella esplorazione di nuovi orizzonti di studio proponendo anche direzioni inedite di ricerca. Da questo punto di vista il lettore potrà notare che lo spirito interdisciplinare delle scienze cognitive è molto adeguato per consentire questa problematizzazione e questo slancio in avanti dell'indagine teorica ed applicativa. Esso crea infatti quella che potremmo definire una peculiare atmosfera di ricerca dentro la quale ci si sente come più liberi rispetto alla tentazione dell'appiattimento in soffocanti strutturazioni capillari della ricerca forzatamente vincolate dalla definizione di snodi e dettagli che sono e spesso rimango-

[1] Independent Scholar, ORCID:0000-0002–4270–1636. E–mail: fnc.research@gmail.com
[2] Università di Trento. E–mail: marco.cruciani@unitn.it
[3] Università di Messina. E–mail: velardi.velardi@gmail.com

no ancora spinosi e problematici, dalla necessità di una previa e aprioristica chiusura sistematica o di una limitante coerenza interna spesso tipica di isolati e minuziosi settori dell'indagine scientifica.

Situarsi all'interno dell'ambito costitutivamente dialogico delle scienze cognitive spinge gli studiosi a far interagire maggiormente le due prospettive della logica (o contesto) della scoperta e della logica (o contesto) della giustificazione spesso pensate in opposizione dalla epistemologia tradizionale, ma che invece possono essere pensate all'interno di un paradigma più interattivo. La distinzione tra i due contesti, elaborata da Hans Reichenbach (1938, 7) risente certamente della radicalità dei modelli del neopositivismo logico che proprio gli studi cognitivi sulla versatilità e fallibilità della concettualizzazione e del ragionamento umano, del *decision making* e del *problem solving*, ci permettono di rimettere in discussione. Qualcuno potrebbe dire che la ricerca cognitiva non fa che dimostrare, al contrario, che il contesto della scoperta può essere oggetto di studio solo da parte della psicologia e che solo il contesto della giustificazione può essere oggetto di studio da parte della epistemologia. A nostro avviso invece il campo delle scienze cognitive mostra come scoperta e giustificazione, psicologia ed espistemologia non siano separabili, ma vadano visti in interazione fra loro, secondo un modello più *transazionale*. Non è solo il contesto *ex post* della giustificazione a potere essere suscettibile di analisi logica e di razionalizzazione da parte della epistemologia, ma vi è tutto il campo della indagine e della ricerca che prepara, in modo spesso non lineare e prevedibile, l'assetto più sistematico della ricerca matura cui poi si può tentare di fornire una veste più deduttiva.

Quest'ultima non è però l'unico modello d'inquadramento di cosa è una scienza, come dimostra an-

che lo studio dei concetti e del ragionamento all'interno di scienze meno deduttive e più probabilistiche come quelle mediche (si veda in questo volume il saggio di Gagliardi). Al contrario, seguendo Charles Peirce, potremmo sostenere, come fatto di recente da Massimo Stevanella (2012) nel suo saggio dedicato non a caso alla teoria dell'abduzione del filosofo americano, che la distinzione di Reichenbach tra contesto della scoperta e contesto della giustificazione non si può sovrapporre a quella tra dominio dell'a–razionalità e dominio della razionalizzazione epistemica. E ci sembra che tutta la ricerca delle scienze cognitive in questo ambito dimostri come l'opposizione neopositivista fosse infondata e scaturisse da un serio pregiudizio nei confronti di un fare scienza meno vincolato alla coerenza deduttiva del sistema, il quale, di par suo, non emerge spontaneamente e linearmente all'interno della ricerca, ma matura attraverso un processo più complesso e spesso più caotico e problematico.

Ci sembra che queste considerazioni caratterizzino bene lo spirito di questa raccolta che, lungi dal descrivere semplicisticamente lo stato dell'arte degli studi, fornisce questa descrizione come qualcosa che si evince dallo sviluppo problematico della ricerca stessa e delle metodologie differenti usati nei diversi settori interdisciplinari.

Il volume che presentiamo al lettore comprende tematiche legate alla robotica, alla teoria della diagnosi, alla filosofia della psichiatria, alle teorie corporee dei concetti, alla trattazione prototipica dei processi di categorizzazione, alla costruzione sociale delle categorie, agli aspetti legati alla concettualizzazione del proprio corpo, alle capacità di apprendimento delle reti neurali artificiali, allo sviluppo della categorizzazione in contesti naturali e artificiali, alla relazione tra rappresentazionea astratta e amodale e *embodiment* sensorimotorio e multimodale, alla relazione tra

concetti e elaborazione del linguaggio soprattutto dal punto di vista della produzione del linguaggio e dell'accesso lessicale.

I contributi spaziano da studi sulle basi teoriche della cognizione a studi basati su applicazioni pratiche, contribuendo alla definizione di un panorama complesso e variegato che inevitabilmente si estende fra discipline differenti, coprendo e intrecciando quasi tutto l'esagono classico delle scienze cognitive: linguistica, filosofia del linguaggio, filosofia della scienza (in particolare della medicina), psicologia cognitiva, neuroscienze, informatica intesa in un senso ampio che comprende l'intelligenza artificiale classica e non, le reti neurali e la robotica. Il panorama delineato condivide un obiettivo comune: la comprensione, la descrizione e la spiegazione dei processi cognitivi di categorizzazione. In questo modo il volume, pur concentrandosi su aspetti specifici della categorizzazionee e della teoria dei concetti, si rivela di grande utilità offrendosi come sintesi peculiare dello stato dell'arte in questo settore della ricerca attraverso una visuale più specialistica che fornisce indirettamente una sorta di teoria generale dei concetti.

Ci siamo permessi di indicare questo rinvio tematico più generale nel saggio introduttivo e riassuntivo, intitolato *Ricerche sulle Teorie dei Concetti e sui Processi di Categorizzazione*, che abbiamo pensato di premettere al volume cercando anche di rintracciare le parentele numerose che legano gli undici saggi che lo compongono. La visione di insieme emerge dagli approfondimenti e non da una previa strutturazione della teoria dei concetti, nelle sue molteplici problematiche e nei suoi molteplici sviluppi, fornendo così una teoria dei concetti già "incarnata" nei diversi scenari concreti di applicazione.

Pensiamo che questa caratteristica possa rendere il volume ancora più interessante e di gradevole lettura sia per coloro che approcciano per la prima volta il complesso tema della teoria dei concetti, sia per coloro che hanno già dimistichezza con questo campo di ricerca e utilizzano il volume come una occasione di ulteriore approfondimento.

Riferimenti bibliografici

REICHENBACH H. (1938) *Experience and prediction. An Analysis of the Foundations and the Structure of Knowledge,* Chicago University Press, Chicago.

STEVANELLA, M. (2012) *La scoperta scientifica e la sua logica,* Mimesis, Milano.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130662
pp. 15-25 (marzo 2018)

Ricerche sulle teorie dei concetti e sui processi di categorizzazione

di Francesco Gagliardi[1],
Marco Cruciani[2], Andrea Velardi[3]

La categorizzazione è un processo cognitivo di suddivisione del mondo in categorie e i concetti sono rappresentazioni mentali delle categorie degli oggetti del mondo. La categorizzazione può essere considerata una condotta adattiva per mezzo della quale gli esseri umani determinano le categorie di oggetti del mondo fisico (ad es. il concetto di uccello), sociale (ad es. il concetto di autorità) e astratto (ad es. il concetto di relazione). In termini adattivi, i concetti possono essere visti come una sorta di 'strumenti mentali' in dotazione agli agenti cognitivi, che legano le esperienze pregresse dell'agente con le sue attuali interazioni con gli oggetti del mondo. In termini epistemologici, i concetti possono essere visti come forme di conoscenza parziale e prospettica con cui gli agenti cognitivi forniscono senso alla realtà (Gagliardi, 2014).

Senza dubbio, la comprensione dei processi di categorizzazione e della natura dei concetti è una sfida intellettuale fra le più dibattute non solo nell'ambito delle scienze cognitive. Essa annovera numerose prospettive e differenti teorie a volte non compatibili fra le quali la teoria classica

[1] Independent Scholar, ORCID:0000–0002–4270–1636. E–mail: fnc.research@gmail.com
[2] Università di Trento. E–mail: marco.cruciani@unitn.it
[3] Università di Messina. E–mail: velardi.velardi@gmail.com

dei concetti, la teoria dei prototipi (Rosch, 1975, 1978, Rosch, Mervis, 1975), la teoria degli esemplari (Medin, Schaffer, 1978), la teoria della teoria dei concetti (Carey 1985, 2009, Gopnik, Meltzoff, 1997, Keil 1989, Medin, 1989, Murphy, Medin, 1985), il c*onceptual atomism* (cfr. Fodor, 1998, Millikan, 2000), la teoria *embodied* e situata dei concetti (Barsalou 2005, Borghi 2006, Clark 1997, Poirer *et al.*, 2005), il *conceptual pluralism* (cfr. Margolis, Laurence, 1999, Weiskopf, 2009) e il *concept eliminativism* (Frixione, 2007, Machery, 2009).

I contributi che seguono indagano la natura dei concetti da varie prospettive e con diverse finalità: la psicologia cognitiva è presente in modo più o meno esplicito nella maggior parte se non in tutti i contributi; la robotica e l'intelligenza artificiale è presente o costituisce lo sfondo dei contributi di Angelo Cangelosi, Antonio Chella, Luciano Celi, Francesco Gagliardi, Stefania Moretti, Alberto Greco, Alessio Plebe e nel secondo di Andrea Velardi, ciò a testimonianza che il computazionalismo (cfr. Cordeschi, Frixione, 2007), nelle sue declinazioni più moderne *embodied* e connessioniste, conferma di essere un utile strumento per indagare i processi cognitivi; le applicazioni in ambito medico sono presenti nei contributi di Romolo Capuano, Francesco Gagliardi ed Elisabetta Lalumera che rappresentano degli interessanti esempi di *cross–fertilization* con la sociologia e la filosofia della medicina; la prospettiva *embodied* e la natura concettuale o meno dei processi di categorizzazione, anche in considerazione del sostrato fisico che realizza le rappresentazioni concettuali, sono presenti, con sfumature diverse e non senza interessanti rapporti dialogici, nei contributi di Edoardo Fugali, Giuliana Gerace, Alessio Plebe e nel primo di Andrea Velardi. Ci sembra utile riportare di seguito una breve introduzione ai singoli contributi raccolti nel volume.

In *Lo Sviluppo dei Concetti nei Robot e nelle Macchine Intelligenti* Angelo Cangelosi e Antonio Chella affrontano il problema dello sviluppo di sistemi intelligenti e robot umanoidi e della capacità da parte delle macchine di percepire il mondo mediante immagini, suoni ed esperienze tattili, di rappresentarlo mediante concetti, e di impiegare questi concetti per svolgere compiti motori, cognitivi e linguistici. In questo articolo gli Autori focalizzano l'attenzione sui modelli computazionali relativi alla capacità dei robot di acquisire i concetti dal mondo esterno, di collegarli alle parole del lessico e di impiegarli per compiti cognitivi in contesti dati (*symbol grounding*). In un modello basato su reti neurali artificiali che prende spunto dall'applicazione alla robotica degli studi sullo sviluppo cognitivo dei bambini verrà analizzato come un robot possa sviluppare autonomamente un lessico e i significati correlati attraverso l'interazione con l'ambiente e mediante strategie motorie. Inoltre, verrà discusso come questo approccio basato su robot possa anche essere esteso all'acquisizione di concetti astratti, come quelli relativi ai numeri.

In *Cleptomania. Genesi ed evoluzione di una categoria criminologica* Romolo Giovanni Capuano ricostruisce la genesi della categoria "cleptomania" dall'Ottocento ad oggi, secondo la prospettiva del costruzionismo sociale, in base alla quale i fenomeni sociali non esistono oggettivamente come dato di natura, ma vengono definiti e costruiti in base a determinati interessi. Capuano analizza la cleptomania nei suoi aspetti storici, criminologici, psichiatrici e morali, ricostruendo l'acceso dibattito intellettuale e disciplinare che, dalla data della sua origine a oggi, si è sviluppato intorno a questo concetto. Attualmente, molti psichiatri esprimono seri dubbi sulla esistenza di una condizione psicologica che possa essere definita scientificamen-

te come "cleptomania" e sulla possibilità di una diagnosi che possa avere una utilità di qualche tipo. Gli stessi dubbi sono condivisi da autorità legali, criminologi, esperti di varia provenienza e opinione pubblica. In sintesi, l'approccio costruzionista consente di assumere un atteggiamento critico nei confronti dei fenomeni sociali, di cui mette in discussione la genesi e le funzioni e, come in questo caso, consente di rivelare i meccanismi sottostanti la costruzione sociale di una categoria.

In *Come le teorie cognitive possono aiutare l'Intelligenza Artificiale*, Luciano Celi propone un'analogia tra alcune caratteristiche delle DNN (*Deep Neural Networks*) e alcune tipiche problematiche delle Scienze Cognitive dove, da tempo, è acclarato che la capacità di categorizzazione tipicamente umana non può essere attribuita al solo aspetto della similarità. L'Autore sostiene che il semplice sottoporre immagini diverse di uno stesso oggetto nella fase di apprendimento non sembra essere garanzia di una valida capacità di categorizzazione sia nell'essere umano che nelle DNN. Almeno un altro elemento sembra necessario: il ricorso alle regole, a una sorta di "lista di controllo" utile a identificare correttamente l'oggetto, soprattutto nella prima fase di apprendimento. La proposta è quindi quella di rivedere anche per le DNN la metodologia di apprendimento della rete stessa non lasciando al solo massivo input di dati il compito di addestrare le DNN, pena i fallimenti che spesso si verificano su compiti semplici, limitando l'uso delle DNN stesse.

In *Concettualizzare l'autoconsapevolezza corporea e le sue basi cognitive. Definizioni e tassonomie* Edoardo Fugali, muovendo dall'assunto che individua nella struttura minimale del *core self* la prima basilare forma di autocoscienza, elabora un'analisi critica dell'apparato concettuale impiegato nelle scienze cognitive per definire

l'autoconsapevolezza corporea. L'Autore propone un approccio che integra le analisi fenomenologiche sull'esperienza del sé corporeo e le indagini sperimentali relative ai meccanismi cognitivi soggiacenti. In particolare, egli prende in esame due specificazioni di questa forma di autocoscienza, ossia il senso di proprietà corporeo e il senso di *agency*, unitamente alle strutture rappresentazionali che li sorreggono – schema corporeo e immagine corporea –, al fine di mostrare come queste categorie corrispondano in sostanza alle dimensioni del corpo vissuto e del corpo oggetto individuate in fenomenologia.

In *I processi diagnostici e le teorie dei concetti* Francesco Gagliardi propone un legame fra le teorie dei concetti e i processi diagnostici. Quest'ultimi, secondo le teorie della diagnosi introdotte in filosofia della medicina, avvengono attraverso due possibili modalità: la diagnosi fisiopatologica e quella nosologica. Lo scopo dell'Autore è mostrare che entrambi si possono considerare come dei particolari processi cognitivi di categorizzazione e concettualizzazione della mente umana effettuati nell'ambito clinico: la diagnosi fisiopatologica è considerata una forma di ragionamento causale *model–based* che rientra nella *theory–theory* dei processi di categorizzazione, mentre la diagnosi nosologica, che si basa sulla similarità, viene considerata come un'attività di categorizzazione con aspetti riconducibili sia alla teoria dei prototipi quanto alla teoria degli esemplari. L'Autore corrobora questa sua analisi anche grazie ad alcuni risultati della psicologia cognitiva, dell'intelligenza artificiale e della filosofia della scienza.

In *Osservazioni sui processi di categorizzazione tra concettualismo e non-concettualismo* Giuliana Gerace fornisce alcune riflessioni sul tema dei processi di categorizzazione e delle relative rappresentazioni cognitive, nel suo

indissolubile legame con il tema della rappresentazione della conoscenza, attraverso argomenti che spaziano dalla psicologia cognitiva alla filosofia della mente. Nella prima parte considera il modo in cui la rappresentazione cognitiva (comunemente definita "concetto") viene trattata all'interno di alcuni fondamentali modelli di categorizzazione teorizzati dalla psicologia: il modello a prototipi e il modello della concettualizzazione situata. Quest'ultimo benché in grado di superare alcuni limiti esplicativi del modello a prototipi si presenta come scarsamente giustificato sul piano teorico proprio con riguardo al tema della rappresentazione percettiva. A questo proposito, la seconda parte de contributo è tesa ad evidenziare come l'indagine filosofica, in particolare quella relativa all'intenzionalismo non–concettualista, abbia prodotto argomenti utili alla giustificazione delle cosiddette rappresentazioni percettive.

In *Concetti e categorizzazione dei disturbi mentali: come la psicologia cognitiva può aiutare la psichiatria* Elisabetta Lalumera mostra che la categorizzazione dei disturbi mentali da parte dei terapeuti non segue una impostazione ateorica e descrittiva come quella del *Manuale diagnostico e statistico dei disturbi mentali* (DSM–5). L'Autrice considera una rassegna di alcuni studi recenti che mostrano una prevalenza di concetti-teoria e di rappresentazioni prototipiche, a seconda del tipo di patologia, e conclude proponendo che gli studi sui concetti usati dei terapeuti possono contribuire a migliorare l'utilità clinica del DSM–5.

In *Categorizzare non vuol dire solo classificare. Alcune riflessioni sui limiti metodologici dell'indagine sperimentale sulla categorizzazione*, Stefania Moretti e Alberto Greco discutono dell'importanza di una teoria della categorizzazione che tenga conto non solo di come vengono

classificati nuovi esempi a partire dalle categorie già apprese, ma anche di come le categorie vengano apprese a partire da esempi. Gli Autori introducono alcuni modelli di psicologia cognitiva mirati alla spiegazione di dati sperimentali ottenuti da compiti di classificazione e cercano di evidenziarne i principali limiti. In particolare, viene analizzato come la portata esplicativa di questi modelli sia limitata per questioni metodologiche. Gli Autori operano un confronto con alcuni metodi in Intelligenza Artificiale per la rappresentazione della conoscenza, per poi proporre un paradigma sperimentale per la formazione di categorie. Concludono discutendo alcune implicazioni teoriche e possibili usi di questo metodo in funzione della varietà e complessità dei processi di categorizzazione.

In *Rappresentazioni corticali* Alessio Plebe sostiene che mentre parlare di rappresentazioni per i neuroni i cui potenziali d'azione provocano una diretta contrazione di qualche muscolo sembra ridondante, risulta invece del tutto naturale quando i neuroni sono coinvolti nel pensare ad oggetti e fatti del mondo. L'Autore affronta la questione a partire dalla connessione tra le nozioni di rappresentazione e di computazione e mostra come, contrariamente a quanto spesso assunto, i supporti teorici al computazionalismo e al rappresentazionalismo sono lontani tra di loro, e come di conseguenza il poderoso e raffinato apparato matematico disponibile riguardo la computazione poco aiuta le rappresentazioni. Plebe suggerisce che per ricavare una nozione formale di rappresentazione un aiuto può venire dalla neuroscienza, cercando di caratterizzare matematicamente la forma che assumono le rappresentazioni nei circuiti neurali, in particolare nella corteccia cerebrale, che è considerata la sede principale di rappresentazioni concettuali, ed è la struttura cerebrale i cui ipotetici principi rap-

presentazionali godono di migliori analisi teoriche e ampi riscontri empirici.

In *Rappresentazione ed embodiment debole. L'intreccio tra astratto e concreto in una ipotesi di doppia elaborazione multimodale e amodale dei concetti* Andrea Velardi affronta il problema del *grounding* discutendo in particolare il ruolo della modalità visiva e dell'immagine mentale come analogato degli esemplari della realtà esterna nel *processing* concettuale. L'Autore propone di integrare la nozione standard di rappresentazione e di *embodiment* attraverso una nozione di rappresentazione più complessa in cui siano rese compatibili la dimensione dell'astrazione e quella del concreto, e che attraverso una nozione di *embodiment debole* preveda l'esistenza di due livelli concettuali: uno multimodale e uno amodale. In questi termini l'Autore ripensa la rappresentazione come qualcosa che si genera attraverso l'attivazione di modalità sensoriali e schemi motori, che poi permettono la costituzione di un dominio di rappresentazione più emancipato e indipendente dal sottostante dominio *embodied*, in cui il concetto può essere trattato di volta in volta in modo più legato alla dimensione concreta degli esemplari o alla dimensione astratta a seconda della complessità della categoria e/o della necessità del contesto, fornendo così sia una base analitica che non analitica alla definizione e comprensione dei concetti: la prima basata più sull'intensione e sulla lista delle caratteristiche, la seconda basata più sul contenuto estensionale relativo agli esemplari concreti.

In *Concetti e produzione del linguaggio. Relazioni categoriali e iperonimia nel modello del lemma* Andrea Velardi analizza il *processing* dei concetti all'interno della produzione del linguaggio e della selezione lessicale focalizzandosi sul modello *WEAVER++* elaborato da Levelt, Roelofs e altri studiosi. In tale modello è centrale lo stadio

del lemma, parola astratta in cui la fase della concettualizzazione non ha ancora portato alla traduzione del concetto lessicale in parola fonica; la relazione tra concettualizzazione e verbalizzazione pone alla rete computazionale il problema della convergenza da parte del parlante verso quell'unica voce lessicale che egli vuole esprimere, verso la selezione del peculiare livello di astrazione pertinente alle loro intenzioni comunicative e al contesto in cui viene enunciato il messaggio. L'Autore affronta il cosiddetto *hypernym problem*, o della ereditarietà dei tratti, per cui le condizioni di un concetto lessicale B (es. leone) attivano anche quelle del suo iperonimo A (es. animale) rendendo compatibile in teoria l'attivazione lessicale di entrambi. Roelofs tende a risolvere il problema mostrando attraverso un modello a diffusione della attivazione in cui la selezione lessicale, ha natura più sintetica, olistica e meno analitico-componenziale della definizione concettuale e del sistema logico della gerarchia dei livelli di astrazione; l'Autore discute una possibile soluzione dell'*hypernym problem* mostrando la complessità della relazione fra i livelli di astrazione e dello *shifting* categoriale tra questi livelli reso possibile nella ideazione lessicale e nell'atto linguistico.

In questo volume il lettore trova una raccolta di argomenti e metodologie relative allo studio sulla natura dei concetti e dei processi di categorizzazione, che pur non potendo essere esaustiva, risulta sufficientemente varia da essere rappresentativa della attività della comunità scientifica italiana e non solo.

Riferimenti bibliografici

BARSALOU, L.W. (2005) *Situated Conceptualization*, in H.

Cohen, C. Lefebvre, (Eds.), *Handbook of categorization in cognitive science*, Elsevier Science, Amsterdam, pp. 619-650.

BORGHI, A. M. (2006) *Vita artificiale e comportamento: simulazioni su categorizzazione e azione*, «Sistemi Intelligenti», 18(1):125–132.

CAREY, S. (1985) *Conceptual Change in Childhood*, MIT Press, Cambridge, MA.

CAREY, S. (2009) *The Origin of Concepts*, Oxford University Press, Oxford.

CLARK, A. (1997) *Being There. Putting Brain, Body, and World Together Again*. MIT Press. Cambridge, MA.

CORDESCHI, R. FRIXIONE, M. (2007) *Computationalism under attack*, in M. Marraffa, M. De Caro e F. Ferretti (eds.) *Cartographies of the Mind: Philosophy and Psychology in Intersection*. Springer, Berlin-Heidelberg, pp. 37–49.

FODOR, J. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press, New York.

FRIXIONE M. (2007), Do concepts exist? A naturalistic point of view. In Penco, C., Beaney, M., Vignolo, M. (a cura di), *Explaining the Mental: Naturalist and Non–Naturalist Approaches to Mental Acts and Processes*, Cambridge Scholars Publishing, Cambridge, UK.

GAGLIARDI, F. (2014) *La naturalizzazione dei concetti: aspetti computazionali e cognitivi*, «Sistemi Intelligenti», 26(2):283–298.

GOPNIK, A., MELTZOFF, A. (1997). *Words, Thoughts, and Theories*, MIT Press, Cambridge, MA.

KEIL, F. (1989) *Concepts, Kinds, and Cognitive Development*, MIT Press. Cambridge, MA.

MACHERY, E. (2009) *Doing Without Concepts*, Oxford University Press, New York.

MARGOLIS, E., LAURENCE, S. (1999) *Concepts: Core*

Readings, MIT Press, Cambridge, MA.

MEDIN, D.L. (1989) *Concepts and conceptual structure*, «American Psychologist», 44(12):1469–1481.

MEDIN, D.L., SCHAFFER, M.M. (1978) *Context theory of classification learning*. «Psychological Review», 85(3): 207–238.

MILLIKAN, R. (2000) *On Clear and Confused Ideas*, Cambridge University Press, Cambridge.

MURPHY, G.L., MEDIN, D.L. (1985) *The role of theories in conceptual coherence*. «Psychological Review», 92(3):289–316.

POIRER, P., HARDY-VALLÉE, B., DE PASQUALE, J.-F. (2005) *Embodied Categorization*, in H. Cohen, C. Lefebvre, (Eds.), *Handbook of categorization in cognitive science*, Elsevier Science, Amsterdam, pp. 739–765.

ROSCH, E. (1975) *Cognitive Representations of Semantic Categories*. «Journal of Experimental Psychology», 104(3): 192–233.

ROSCH, E. (1978) *Principles of Categorization*, in E. Rosch & B. B. Lloyd (eds.), *Cognition and Categorization*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 27–48.

ROSCH, E., MERVIS, C. (1975) *Family Resemblances: Studies in the Internal Structure of Categories*, «Cognitive Psychology», 7: 573–605.

WEISKOPF, D. (2009) *The Plurality of Concepts*. «Synthese», 169: 145–173.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130663
pp. 27-56 (marzo 2018)

Lo sviluppo dei concetti nei robot e nelle macchine intelligenti

di Angelo Cangelosi[1],
Antonio Chella[2]

1. Introduzione

Lo sviluppo di sistemi intelligenti interattivi come i robot umanoidi o mobili, o le macchine intelligenti come ad esempio un'automobile autonoma o un assistente digitale, richiede la capacità da parte delle macchine di percepire il mondo esterno mediante immagini, suoni ed esperienze tattili, di rappresentarlo mediante concetti, di utilizzare questi concetti per compiti motori e cognitivi, e per comunicare mediante il linguaggio con altri agenti (persone e altre macchine). In questo capitolo ci focalizzeremo sulle metodologie e modelli computazionali relativi alla capacità dei robot e delle macchine intelligenti (che possiamo chiamare in maniera più generale agenti cognitivi) di acquisire i concetti dal mondo esterno, di collegarli alle parole del lessico e di impiegarli per compiti cognitivi e linguistici. Per esempio, in una visione futura di un robot che assiste gli anziani in una casa di riposo aiutando il personale infermieristico nel distribuire il cibo, il robot deve sapere che la percezione visiva (e olfattiva e tattile) di diver-

[1] Plymouth University e Manchester University, UK, E–mail: a.cangelosi@plymouth.ac.uk
[2] Università degli Studi di Palermo e ICAR–CNR, Palermo, Italia

se mele, o della stessa mela da diversi punti di vista, fa parte dello stesso concetto di <mela> (un frutto edibile, rotondo, di colore tipicamente rosso e giallo), e che la parola "mela" è utilizzata per descrivere questa categoria di frutta[3]. Di seguito vedremo come gli studi allo stato dell'arte sulla formazione dei concetti e delle parole per descrivere gli stessi possano essere usati con i robot.

Un robot deve essere in grado di generare il concetto di "mela" partendo dalla percezione visiva di diverse immagini del frutto fisicamente presente nell'ambiente (lo stesso principio si applica per la percezione tattile e olfattiva) e dalla formazione di una rappresentazione interna concettuale del frutto, come nel caso della percezione delle categorie di cui parleremo più avanti. Questo processo è descritto come *perceptual anchoring* (Coradeschi, Saffiotti, 2003). Il processo più generale di collegamento delle parole con le entità del mondo esterno, che ne costituiscono il significato, è chiamato *grounding* (Harnad, 1990; Cangelosi, 2010). Più specificamente, il problema di collegare intrinsecamente i simboli usati da un agente cognitivo ai loro corrispondenti significati e alle entità del mondo esterno è chiamato *symbol grounding problem* (Harnad, 1990; Cangelosi *et al.*, 2000). Perché un agente cognitivo sia considerato un modello psicologicamente plausibile delle categorie concettuali e linguistiche umane, i simboli (parole) devono essere intrinsecamente legati alla capacità dell'agente di acquisire le corrispondenti categorie mediante l'interazione con l'ambiente e senza la mediazione di una persona (o di un programmatore) che stabilisce esplicitamente il collegamento tra la parola ed il concetto. In

[3] In questo capitolo per convenzione utilizzeremo la notazione <concetto> per riferirsi a uno specifico concetto, e "parola" per riferirci alla parola usata per nominare il concetto.

particolare, è essenziale che alcuni simboli di base siano direttamente collegati (*grounded*) alle categorie sensomotorie. Successivamente, nuove parole e i loro significati e relativi concetti, possono essere formate attraverso meccanismi di *grounding* indiretto (*grounding transfer*) basati sulla combinazione di simboli e categorie di base. Per esempio, posso inventare il concetto e la parola di "unicorno" combinando i concetti collegate alle parole per "cavallo" e "corno".

Nelle sezioni successive verranno prima analizzati i diversi approcci computazionali alla rappresentazione e apprendimento di categorie in sistemi robotici e artificiali (Sezione 2). Questa sezione include una discussione dei metodi di rappresentazione simbolica dei concetti in approcci di intelligenza artificiale classica, seguita dall'analisi degli approcci di apprendimento e rappresentazioni distribuite basati sulle reti neurali, e infine quelli del metodo degli spazi concettuali. Verranno poi presentati in dettaglio due esempi di uso di concetti artificiali in robot e agenti artificiali. Il primo riguarda l'apprendimento del linguaggio attraverso il grounding delle parole in concetti di esperienze percettive di un robot umanoide (Sezione 3). Il secondo esempio si focalizza sull'uso di rappresentazioni concettuali per la percezione del sé in agenti artificiali simulati (Sezione 4). Questi esempi servono a mostrate il potenziale dell'uso di modelli computazionali dei concetti in varie capacità cognitive, così come il loro utilizzo in diverse tipologie di agenti artificiali robotici e simulati.

2. Rappresentazione di concetti nei robot e macchine intelligenti

Nella letteratura sulla rappresentazione dei concetti negli agenti artificiali, sono stati analizzati e messi a confronto essenzialmente tre tipi di rappresentazione (cf. Figura 1): (i) le rappresentazioni simboliche, (ii) le rappresentazioni neurali, e (iii) il livello intermedio degli spazi concettuali. Nelle rappresentazioni simboliche, i concetti sono definiti mediante le loro caratteristiche e le loro relazioni con altri concetti e strutture basate sulla manipolazione di simboli e sull'uso della logica, come nelle reti semantiche, di cui tratteremo nel dettaglio più avanti, basate su sistemi di conoscenza simbolica definita dall'utente umano. Nelle rappresentazioni neurali un concetto è definito mediante schemi di attivazione, tipicamente distribuiti e paralleli, di neuroni simulati nei sistemi di reti neurali artificiali, ed è appreso durante addestramento. Tra i due livelli possiamo considerare un livello intermedio di rappresentazione dei concetti basato sugli spazi concettuali, introdotti da Peter Gärdenfors (2000; 2014). Un concetto è rappresentato mediante un insieme di punti dello spazio concettuale, ossia uno spazio metrico caratterizzato da dimensioni qualitative, come ad esempio la forma, le componenti di colore, l'altezza di un suono, il volume. Questi tre tipi di rappresentazione verranno discussi nel dettaglio nei paragrafi successivi.

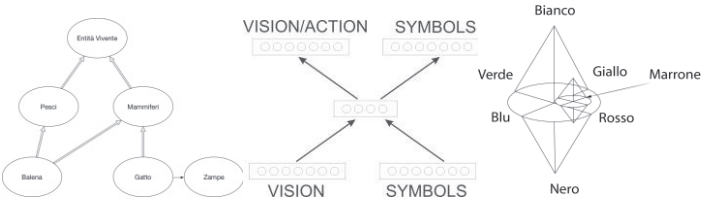

Figura 1. Esempi di reti semantiche (sinistra), reti neurali a doppia via (centro), e spazi concettuali (destra).

2.1 *Concetti mediante rappresentazioni simboliche*

L'approccio simbolico-cognitivista usa rappresentazioni simboliche ed è basato sulle teorie psicologiche di Elaborazione Umana dell'Informazione (Human Information Processing: Newell, Simon, 1972), tipiche delle prime metodologie di Intelligenza Artificiale. Secondo tali teorie, il sistema cognitivo umano ha una architettura simile a quella dei calcolatori elettronici, e i processi cognitivi sono processi basati sulla manipolazione di rappresentazioni simboliche. Questa visione della mente come un computer implica che i processi sono sequenziali (cioè delle regole che vengono applicate in sequenza, una dopo l'altra), e che i diagrammi di "blocchi-e-frecce" tipiche delle rappresentazioni di algoritmi possono essere usate anche per rappresentare funzioni cognitive. Un tipico esempio è quello delle regole per la formazione del tempo passato (*past tense*) dei verbi in inglese, che ha due "blocchi" rappresentanti due diverse strategie sintattiche: il blocco regolare riguarda l'uso della regola "aggiungere -ed" per i verbi regolari; l'altro blocco contiene la lista di verbi con morfologia irregolare, come "go→went; be→was." Le frecce collegano questi blocchi a seconda che il verbo abbia struttura regolare o irregolare.

L'approccio simbolico–cognitivista all'origine dei metodi computazionali per la simulazione della mente, sostiene che l'attività mentale si riduce alla rappresentazione del mondo mediante simboli e regole logiche di manipolazione dei simboli, come nel classico esempio della regola: SE–ALLORA (*IF–THEN*). Quindi, se voglio simulare le capacità concettuali e linguistiche di un agente cognitivo, nel mio programma devo elencare i simboli (chiamate variabili, nei programmi per computer) che rappresentano i concetti (ad esempio <cane>, <gatto>). Questi simboli concettuali possono a loro volta essere collegati ad altri simboli, per esempio il concetto <cane> avrà come proprietà quella di avere quattro zampe, una coda, di scodinzolare, abbaiare e così via. Successivamente, userò simboli per identificare le parole che mi permettono di parlare di questi concetti (ad es. "cane"), e variabili simboliche per le rappresentazioni fonetiche (ad esempio "cane" è basato sui fonemi, anch'essi simbolici, come /c/ /a/ /n/ /e/). Inoltre, è possibile usare rappresentazioni simboliche per la manipolazione di simboli mediante le regole sintattiche. Ad esempio, la regola "Soggetto–Verbo–Oggetto" per le costruzioni di verbi transitivi descrive la struttura in cui il "Soggetto" dell'azione deve sempre precedere il "Verbo", seguito dal complemento "Oggetto." Un altro esempio è quello delle regole logiche di deduzione delle conoscenze, come la regola "SE <cane> ALLORA <abbaia>". In questo modo, se penso al significato del concetto di cane allora posso dedurre un concetto motorio ad esso collegato, quale è l'azione di abbaiare. L'approccio simbolico è usato nelle basi di conoscenza e nelle ontologie dei sistemi esperti, ma anche nei sistemi intelligenti di pianificazione e ragionamento.

Un altro approccio simbolico per la rappresentazione delle categorie, in particolare per le strutture semantiche

gerarchiche, è basato sulle reti semantiche (si veda ad esempio Sowa, 1992). Un classico esempio di rete semantica è quello del regno animale (Figura 1, sinistra), che collega la gerarchia di categorie dal nodo/concetto superiore (<entità vivente>), alle sue sotto categorie come <pesci> e <mammiferi>, e successive sotto–categorie, come il <gatto> e la <balena> per i mammiferi e così via. I collegamenti tra questi concetti gerarchici sono chiamati *IS–A*, traducibile in "È un" o "È un tipo di" (cioè <gatto> IS–A <mammifero>). Inoltre, ogni nodo concettuale ha anche altri tipi di collegamenti riferiti alle sue proprietà, come *HAS,* per "ha" (cioè <gatto> HAS <zampe>). In questo tipo di rappresentazione, la struttura dei concetti è determinata a priori dal programmatore.

Una importante caratteristica dei metodi simbolico–cognitivisti, e allo stesso tempo un limite importante per lo studio della mente, è la mancanza dei processi di apprendimento. Questi modelli in genere implementano direttamente la conoscenza (basata su simboli e regole) di un adulto, piuttosto che considerare un algoritmo che cambi queste rappresentazioni e regole durante l'apprendimento mediante l'interazione con il mondo esterno e con altri agenti.

2.2 *Concetti mediante reti neurali artificiali*

L'approccio basato sulle reti neurali artificiali, chiamato anche approccio connessionista, è basato sulle teorie della mente inspirate dal funzionamento del cervello mediante reti di neuroni, e utilizza una architettura di Elaborazione Parallela e Distribuita (*Parallel Distributed Processing*: Rumelhart, McClelland, 1986; Parisi, 1989). Queste teorie sono basate sui modelli simulativi delle reti neurali artifi-

ciali, dette reti connessioniste. Ogni funzione cognitiva è descritta in base ai principi funzionali e strutturali del cervello, con l'uso di popolazioni di neuroni che possono attivarsi in parallelo per rappresentare i concetti e le conoscenze in maniera distribuita e in diversi circuiti neurali. Una tipica rete neurale ha uno strato di unità di ingresso (che rappresentano i sensori visivi, uditivi, tattili) che è connesso ad uno strato di unità intermedie (unità nascoste, chiamate *hidden*) che è infine connesso allo strato di unità di uscita (che rappresentano le unità motorie per il controllo di azioni, o il concetto a cui corrisponde l'immagine in ingresso). Le architetture neurali più complesse includono anche le connessioni ricorrenti e gli strati di unità di memoria per implementare la capacità di ricordare gli eventi passati e per le opportune strategie cognitive ricorsive. In queste architetture di ispirazione neurale non ci sono simboli o regole esplicite. Al contrario, i concetti, le parole, e i significati sono rappresentati dalle unità nascoste in maniera parallela e distribuita, e per questo i sistemi connessionisti sono anche chiamati subsimbolici (*subsymbolic*). Quindi, se in ingresso mostro una immagine di un cane, le attivazioni dello strato nascosto costituiscono delle rappresentazioni concettuali distribuite, cioè costituite dai valori di attivazione di tutte le unità, invece che dalla risposta una singola unità. Nei più recenti modelli di apprendimento profondo (*deep learning)*, come le reti neurali convoluzionali (LeCun *et al.*, 2015), queste rappresentazioni intermedie sono chiamate *features* e sono man mano più compatte passando dallo strato di ingresso verso i livelli successivi di strati nascosti.

Le reti neurali connessioniste hanno l'importante caratteristica di implementare metodi di apprendimento che modificano i pesi delle connessioni tra neuroni (sinapsi) in risposta alle fasi di addestramento. Per esempio,

l'algoritmo di apprendimento associativo detto *hebbiano*, basato sulla teoria neuronale di Donald O. Hebb (1949), (che è stata dimostrata essere effettivamente presente nel cervello, dopo la scoperta del meccanismo LTP — Long Term Potentiation), incrementa il peso delle connessioni tra due neuroni quando questi sono simultaneamente attivi. Un altro algoritmo di grande diffusione nei modelli connessionisti è quello detto della retropropagazione (*back-propagation*), che implementa un sistema di apprendimento supervisionato dove l'istruttore (genitore, insegnante) fornisce all'allievo la risposta corretta, che, confrontata con quella suggerita dall'allievo, produce una differenza (errore) che a sua volta viene propagata all'indietro nelle connessioni della rete. Questi metodi di apprendimento permettono di simulare il processo dello sviluppo di apprendimento del bambino, e consentono l'analisi delle prestazioni di un agente cognitivo durante le diverse fasi del suo sviluppo.

Tipicamente, i modelli di reti neurali per il *grounding* dei concetti e delle parole utilizzano un'architettura "a due vie" (Figura 1, centro). L'architettura a due vie coinvolge sia l'ingresso visivo (ad esempio la proiezione sulla retina o l'elenco delle funzioni visive) che l'ingresso linguistico (ad esempio la codifica locale o grafemica/fonetica dei simboli). Lo strato di uscita avrà opportune unità simboliche per la rappresentazione di nomi e parole (ad esempio con una codifica fonetica degli elementi lessicali) e una rappresentazione categorica degli stimoli di ingresso (ad esempio un nodo per ogni categoria o una rappresentazione visiva dei prototipi di categoria). Alcuni modelli possono utilizzare una rappresentazione delle categorie basata sulle azioni, anche se questo metodo è tipico dei modelli connessionisti negli agenti robotici incarnati (*embodied*). Tutti gli strati di ingresso e di uscita sono collegati tramite

unità nascoste. Il percorso dall'ingresso visivo all'uscita simbolica è utilizzato per l'attività di produzione linguistica, come la denominazione dell'oggetto rappresentato nella scena visiva e della sua categoria. Questo è il percorso essenziale di una rete per il *grounding* di simboli, in quanto il collegamento visione–linguaggio costituisce il meccanismo fondamentale del grounding percettivo. Il percorso dall'ingresso linguistico all'uscita visiva / categoriale è utilizzato per le attività di comprensione del linguaggio.

Uno dei primi e più influenti modelli di denominazione e di *grounding* è stato sviluppato da Plunkett e collaboratori (Plunkett *et al.*, 1992, Plunkett, Sinha, 2011). L'architettura neurale si basa sulla rete standard a due vie. Il risultato più interessante è che le prestazioni di apprendimento non sono lineari, ma attraversano fasi di sviluppo a stadi, cioè improvvisi miglioramenti dopo periodi di stasi, come nel caso dell'effetto detto *vocabulary spurt* (Plunkett *et al.*, 1992). La simulazione riflette ciò che è effettivamente osservato nei bambini o negli adulti quando si apprende una nuova lingua: la comprensione precede la produzione.

I modelli connessionisti sviluppati da Harnad *et al.* (1991; 1995) si sono concentrati sul fenomeno del *grounding* di simboli e della percezione categoriale. Il compito di queste reti neurali consiste nel classificare delle linee in base alla loro lunghezza. Tali linee sono state rappresentate da unità di ingresso che usano due schemi di codifica: la codifica iconica (ad esempio una linea di lunghezza 4 potrebbe essere codificata come "11110000") o la codifica posizionale (ad esempio, la stessa linea è codificata come "00010000"). L'apprendimento consiste in due attività sequenziali: l'autoassociazione e l'apprendimento della categoria. Il primo compito permette alla rete di "discriminare" tra diversi stimoli utilizzando una pre–categorizzazione mediante l'apprendimento auto–

associativo. I vettori di attivazione delle unità nascoste sono esaminati per registrare le distanze di percezione categoriale per ogni coppia di ingresso. Dopo l'attività di autoassociazione, le reti sono addestrate per categorizzare gli stimoli in tre categorie: breve, medio, lungo. Il confronto tra distanze e post–categorizzazione nelle attivazioni delle unità nascoste delle reti ha mostrato l'effetto ottico della compressione delle distanze tra gli stimoli della stessa categoria, e l'incremento della distanza tra stimoli appartenenti a categorie diverse. Harnad *et al.* (1991; si veda anche Tijsseling, Harnad, 1997) hanno anche scoperto che le distanze tra le attivazioni di unità nascoste sono già massimizzate durante l'auto–associazione, per effetto del compito di discriminazione. Questa separazione, tuttavia, non è sempre così chiara e tale da consentire la separabilità lineare nell'iperspazio delle attivazioni delle unità nascoste, come avviene nel caso di stimoli perfettamente classificati. L'algoritmo di *backpropagation*, che simula l'apprendimento delle categorie attraverso una retroazione controllata, ha l'effetto di aggiustare tali rappresentazioni poco chiare e di formare un iperpiano che separa i membri di diverse categorie. Il risultato è una migliore organizzazione delle rappresentazioni categoriali.

Un diverso tipo di rete neurale usata per compiti di categorizzazione è quello delle mappe auto-organizzanti (*Self–Organizing Maps*) dette anche reti di Kohonen (Kohonen 2001) Queste reti sono a due strati, dove l'ingresso rappresenta un'immagine (tipicamente elaborata per ridurre il numero di unità) e lo strato di uscita è organizzato in due dimensioni (chiamato appunto mappa). L'apprendimento avviene in modo non–supervisionato, cioè la rete scopre la similarità tra gli stimoli di ingresso attraverso opportuni metodi di modifica dei pesi mediante un algoritmo competitivo. Alla fine dell'addestramento,

ogni stimolo tipicamente attiva una sola unità che lo rappresenta (chiamata *winner–take–all*) e gli stimoli si auto–organizzano opportunamente per formare una mappa somatotopica (come nella corteccia visiva o motoria) dove gli stimoli molto simili attivano unità vicine, mentre gli stimoli molto diversi attivano unità distanti nella mappa. Come vedremo successivamente, questo tipo di reti può essere utilizzato per l'apprendimento delle parole nei robot.

2.3 *Concetti mediante spazi concettuali*

Gli spazi concettuali consentono una rappresentazione dei concetti intermedia, di tipo geometrico, tra le reti neurali e le rappresentazioni simboliche.

Uno spazio concettuale è uno spazio metrico in cui ogni entità è caratterizzata da un insieme di dimensioni qualitative (Gärdenfors, 2000; 2014). Esempi di dimensioni sono il colore, l'altezza del suono, il volume, le coordinate spaziali e così via. Alcune dimensioni sono strettamente correlate agli ingressi sensoriali, altre possono essere caratterizzate in termini più astratti.

Ad esempio, lo spazio concettuale relativo al dominio della percezione visiva sarà caratterizzato dalle dimensioni relative al colore, alla forma, alla posizione spaziale dell'oggetto percepito. Una specifica <mela> corrisponderà quindi ad un punto specifico dello spazio concettuale corrispondente al colore, alla forma e posizione della mela percepita.

Le dimensioni di uno spazio concettuale rappresentano le qualità di una entità indipendentemente da qualsiasi descrizione di tipo simbolico. Le dimensioni di uno spazio concettuale sono ad un livello di astrazione più generale

degli schemi di attivazione delle reti neurali; in questo senso la descrizione di un concetto mediante le dimensioni di uno spazio concettuale è ad un livello di astrazione intermedio tra il livello simbolico e il livello connessionista.

Un importante aspetto della teoria degli spazi concettuali è la definizione della funzione di metrica dello spazio. Secondo Gärdenfors (ibidem) la distanza tra due punti di uno spazio concettuale calcolata in base a tale metrica corrisponderà alla somiglianza fra le corrispondenti entità. Così ad esempio in uno spazio concettuale che rappresenta la frutta, il punto corrispondente alla "mela annurca" sarà più vicino al punto corrispondente alla "mela cotogna" rispetto al punto corrispondente ad una "pera." In questo spazio, il concetto di <mela> è rappresentato dall'insieme di punti corrispondenti alle mele.

Gärdenfors introduce il cosiddetto *Criterio P*, secondo cui le categorie *naturali* corrispondono a particolari tipi di insiemi, ossia agli insiemi convessi. La caratteristica di tali insiemi, lo ricordiamo, è che, comunque presi due punti appartenenti all'insieme, tutti i punti compresi tra questi due appartengono ancora all'insieme e ne condividono le caratteristiche. Così se prendiamo tutti i punti compresi tra la "mela annurca" e la "mela cotogna," questi corrispondono ancora a delle mele e fanno parte del concetto di <mela>. Secondo Rosch (1975), le categorie naturali rappresentano il livello più informativo di categorizzazione nelle tassonomie delle entità del mondo reale. Sono le più differenziate l'una dall'altra e costituiscono il livello preferenziale di riferimento. Inoltre, sono le prime ad essere apprese dai bambini, e la loro individuazione è più veloce.

Un importante aspetto degli spazi concettuali è la possibilità di rappresentare il *prototipo*: il prototipo di un insieme convesso è il punto centrale dell'insieme stesso, e rappresenta il singolo elemento più caratterizzante del

concetto. Ad esempio, se consideriamo l'insieme degli uccelli, l'elemento più centrale dell'insieme sarà il prototipo di <uccello>, e sarà presumibilmente simile ad un passero e non, mettiamo, ad un pinguino. Infatti, parlando di un uccello è più comune immaginare un passero che non un pinguino.

La Figura 1 (destra) rappresenta lo spazio concettuale relativo ai colori, rielaborata da Gärdenfors (ibidem), che prende la forma della trottola dei colori. Ogni punto rappresenta un colore, e colori simili sono rappresentati da punti vicini nello spazio. Nella figura si può notare una trottola più piccola, ossia un insieme di punti dello spazio concettuale, corrispondente alle gradazioni del colore marrone.

3. Uso di concetti nei robot intelligenti

Di seguito descriviamo due esempi di agenti cognitivi che utilizzano rispettivamente sistemi neurali e metodologie di robotica dello sviluppo per acquisizione dei primi concetti e parole nei robot, e l'uso degli spazi concettuali per la percezione del mondo esterno e del sé in un robot.

3.1 *Acquisizione delle prime parole in robot bambini*

Mettendo insieme metodi di categorizzazione mediante reti neurali e metodi di robotica dello sviluppo, che sono ispirati alla psicologia dello sviluppo, è possibile creare un modello cognitivamente plausibile dell'apprendimento dei concetti e delle parole nei robot, come descritto da Morse *et al.*, (2015).

La robotica dello sviluppo (*developmental robotics*; Cangelosi, Schlesinger, 2015) è l'approccio interdisciplinare alla generazione delle capacità comportamentali e cognitive in un robot, che prende spunto dai principi e dai meccanismi dello sviluppo osservati nei bambini, tramite esperimenti di psicologia dello sviluppo. In particolare, l'idea principale è che il robot, utilizzando una serie di principi intrinseci che regolano l'interazione in tempo reale tra il corpo, il cervello e il suo ambiente, può autonomamente acquisire un insieme sempre più complesso di capacità mentali e sensomotorie. Questa idea è alla base del modello di acquisizione di concetti e parole del robot umanoide bambino *iCub*, e prende ispirazione da esperimenti di psicologia dello sviluppo secondo cui la postura del corpo, e quindi le strategie di *embodiment*, svolgono un ruolo fondamentale nell'apprendimento delle prime parole in un bambino. Samuelson *et al.*, (2011) hanno mostrato che un bambino che fa esperienza di due nuovi oggetti (cioè l'oggetto bersaglio da denominare, e un oggetto distrattore), mostrati ripetutamente in posizioni costanti ma diverse (rispettivamente alla sinistra e alla destra su un tavolo), quando sente per la prima volta il nome dell'oggetto bersaglio (la parola artificiale "modi") con gli oggetti nascosti, ma allo stesso tempo con l'attenzione rivolta verso la posizione dell'oggetto bersaglio (sinistra), il bambino è in grado di imparare a riconoscere come "modi" l'oggetto che occupava la stessa posizione quando è stato nominato. Questo significa che i bambini usano una strategia basata sulla memoria della propria postura e della posizione relativa dell'oggetto per associare gli oggetti ad i loro nomi. Questo esperimento è stato adattato ad un modello di apprendimento del linguaggio mediante il ruolo della postura per il robot *iCub*. Morse *et al.* (2010; 2015) hanno sviluppato una architettura di robotica dello sviluppo, chiamata

epigenetica, basata su una serie di mappe auto-organizzanti di Kohonen e sull'apprendimento *hebbiano* precedentemente descritto. Il nucleo di tale architettura è costituito da tre mappe per le rappresentazioni concettuali. La prima mappa (visiva) è stata pre–addestrata mediante l'algoritmo competitivo di Kohonen, per la categorizzazione di proprietà visive (per esempio utilizzando in ingresso lo spettrogramma HSV del colore di ciascun oggetto al centro dell'immagine). La seconda mappa (postura) riceve in ingresso le informazioni sulla postura del robot (i valori attuali degli occhi, e del torso del robot) e attiva in uscita una mappa somatotopica simile (ma molto semplificata) all'homunculus corticale. La terza mappa (le parole) risponde in modo univoco ad ogni parola sentita durante l'interazione (parola foneticamente riconosciuta dal software di riconoscimento del parlato di Dragon Dictate™). La mappa dei colori e la mappa delle parole sono entrambe completamente collegate alla mappa della postura, con pesi di connessione regolati dalla regola di apprendimento *hebbiano*. Le unità all'interno di ogni mappa sono inoltre completamente collegate all'interno di ciascuna mappa con connessioni inibitorie fisse, per simulare la struttura dei modelli di attivazione interattiva e competitiva (McClelland, Rumelhart, 1981). Il comportamento iniziale, istintivo del robot *iCub* è guidato dalla capacità e dalla curiosità di guardare gli oggetti in movimento (mediante un algoritmo basato sulle mappe di salienza visiva che rileva quali oggetti, o parti del corpo, si siano mosse).

In una ulteriore versione dell'esperimento, l'oggetto bersaglio (una pallina rossa) viene posizionato a sinistra dell'*iCub*. Il robot esamina il bersaglio per circa 10 secondi, prima che l'oggetto venga rimosso e l'oggetto distrattore sia posizionato a destra dell'*iCub*, così che il robot lo guardi. Lo sperimentatore muove gli oggetti per attivare la

mappa di salienza motoria e far girare l'*iCub* (cioè fa cambiare la sua postura) per fissare l'oggetto nella parte sinistra o destra del tavolo. Questa procedura viene ripetuta quattro volte. Al quinto ciclo di presentazione, l'oggetto distrattore viene collocato nella posizione normalmente associata all'oggetto bersaglio e viene detta la parola "modi." Le posizioni originali di ciascun oggetto vengono ripetute per un'ultima volta, e poi nella fase finale di test entrambi gli oggetti sono messi in una nuova posizione del tavolo (al centro) chiedendo al robot "trova i modi". L'*iCub* fissa uno dei due oggetti, quello che ha associato indirettamente alla parola "modi." Sono state eseguite varie versioni dell'esperimento, ognuna ripetuta venti volte (con tutti i pesi di apprendimento ri–inizializzati in maniera casuale, e contro–bilanciando le posizioni di sinistro–destra per il bersaglio e distrattore). I dati degli esperimenti robotici mostrano gli stessi risultati statistici dei dati sui bambini. Per esempio, se la posizione sinistra/destra nell'esperimento di base è tenuta costante rispettivamente per il bersaglio/distrattore, il robot, come i bambini, associa la parola "modi" all'oggetto mostrato a sinistra. Ma se la posizione viene scambiata il 50% delle volte, sia il robot che i bambini fanno scelte causali nel test finale. Un'estensione interessante di questo modello robotico consiste nell'aggiungere ulteriori variazioni dell'esperimento, prima che queste fossero ripetute in esperimenti con bambini. Oltre a variare la posizione orizzontale degli oggetti, e conseguentemente la postura di sinistra/destra, è stata aggiunta una posizione verticale alto/basso (il tavolo è tenuto in due posizioni di altezza, che costringono il robot a muovere il suo torso e la sua postura). I dati robotici hanno mostrato un effetto inaspettato. Cioè quando l'*iCub* sta in posizione alta e l'oggetto bersaglio viene spostato da sinistra a destra, il robot continua a

chiamare "modi" l'oggetto bersaglio anche (mentre ci si sarebbe aspettati un effetto di interferenza tra i due compiti, visto il bersaglio è stato spostato a destra). Successivi esperimenti con bambini hanno introdotto questa nuova variazione alto/basso, mostrando che l'interferenza scompare anche con i bambini. Quindi il modello cognitivo dell'*iCub* è stato in grado di prevedere nuovi meccanismi di acquisizione del linguaggio nei bambini. Questo mostra l'importanza di una collaborazione diretta tra la robotica e la psicologia dello sviluppo.

La stessa architettura cognitiva è stata usata, sempre in collaborazione con gli psicologi dello sviluppo, per simulare il meccanismo della mutua esclusione (*mutual exclusivity,* Twomey *et al.* 2016). In questo caso, si consideri la situazione in cui il bambino vede due oggetti, un trenino e un fischietto. Il bambino ha visto tante volte il trenino e ne conosce il nome, ma non sa il nome del fischietto. Quando lo sperimentatore dice "guarda il fischietto" il bambino è in grado di capire, per mutua esclusione, che "fischietto" è il nome dell'oggetto/categoria <fischietto>. Gli esperimenti con l'*iCub* di Twomey *et al.* (2016) hanno replicato esattamente i risultati di esperimenti di mutua esclusione con bambini.

Un'estensione di questo approccio della robotica dello sviluppo è stato applicato anche ai concetti astratti. Per esempio, Stramandinoli *et al.* (2017) hanno simulato l'acquisizione di concetti e parole astratte come "usare", per esempio "usare un martello", "usare una sega", ma anche potenzialmente per un livello ancora più astratto come "usare un'idea". Altri hanno studiato il *grounding* dei concetti numerici nei robot, usando strategie di *embodiment* come i gesti del contare (Rucinski *et al.* 2012) o il contare con le dita della mano (De La Cruz *et al.* 2014).

3.2 *Spazi concettuali per rappresentazione del sé*

In questo paragrafo analizziamo l'uso degli spazi concettuali nel sistema di visione di un robot in grado di utilizzare concetti legati alla percezione del mondo esterno. Inoltre, introduciamo alcune problematiche relative alla rappresentazione dei concetti legati al senso di sé, ossia la percezione di un robot che riflette su se stesso.

Questo paragrafo ha un doppio fine: far vedere come, tramite la teoria degli spazi concettuali, un robot possa effettuare il *grounding* dei suoi propri concetti mediante rappresentazioni interne al robot stesso e inoltre mostrare come il *grounding* possa essere effettuato anche per i concetti legati al ragionamento introspettivo del robot.

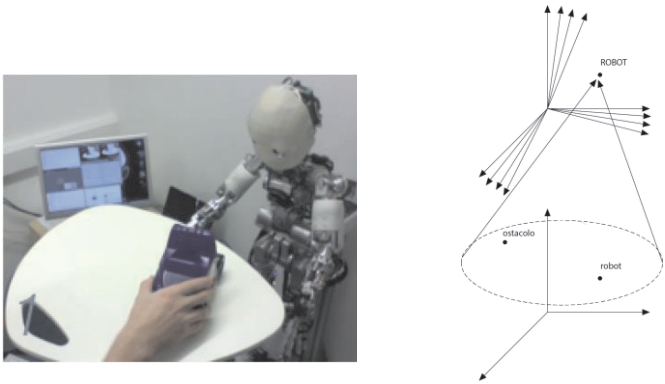

Figura 2. Setup per esperimento con il robot *iCub* sulle parole (sinistra) e modello di rappresentazione del sé (destra).

Chella *et al.* (1997) descrivono diversi esperimenti relativi al sistema di visione di un robot equipaggiato con gli spazi concettuali. Il robot è in grado di descrivere semplici scene statiche costituite da palle, martelli, chiodi: il robot dice che nella scena c'è una "palla", un "martello" e che la

palla è "a destra del" martello, che il martello "è fatto da" una testa e un manico, e così via. Attraverso gli spazi concettuali il robot è in grado di descrivere non solo gli oggetti presenti, ma anche le relazioni tra essi, corrispondenti a opportune relazioni nello spazio concettuale. Il robot è inoltre in grado di generare aspettative sulla scena, quindi, ad esempio, quando il robot riconosce qualcosa che assomiglia al manico di un martello, allora esplora le sue vicinanze per confermare o rigettare l'ipotesi che un martello sia effettivamente presente nella scena. Inoltre, se il robot percepisce durante il suo funzionamento diverse scene in cui un martello è sempre presente insieme ad una palla, allora, quando rivede il martello, genera l'aspettativa della presenza della palla. Quindi, il robot è in grado di descrivere la scena in maniera ricca mediante una esplorazione opportunamente guidata dagli spazi concettuali.

Il sistema di visione del robot è stato esteso (Chella *et al.*, 2000) considerando anche il *tempo* come ulteriore dimensione dello spazio concettuale. Il robot è quindi in grado di percepire le scene in movimento e di utilizzare i concetti legati alle azioni compiute dalle persone. Il robot riconosce ad esempio l'azione di una persona che passa un oggetto ad un'altra persona, riconoscendo che la persona sta allungando il braccio per prendere l'oggetto, che lo prende e che lo porge all'altra persona, e che questa per prenderlo a sua volta estende il proprio braccio. Lo spazio concettuale del robot si arricchisce di concetti legati alle azioni, come "prendere", "lasciare", "passare", "estendere".

Lo spazio concettuale può essere descritto dal programmatore, ma può anche essere appreso dal robot stesso. Una metodologia di apprendimento dei concetti nello spazio concettale va sotto il nome di apprendimento per imitazione (Schaal 1999; Schaal *et al.* 2003). Secondo

questa metodologia esplorata in Chella *et al.* (2006), il robot apprende i concetti "metti a destra di", "metti a sinistra di", guardando un istruttore e cercando di imitarlo.

Lo spazio concettuale è a tutti gli effetti una struttura simulativa, nel senso che il robot può utilizzare lo spazio concettuale non solo per descrivere le scene effettivamente percepite dai sensori, ma può descrivere scene immaginate mediante una *simulazione* (Hesslow, 2002; Grush, 2004). Chella, Macaluso (2009) hanno utilizzato questa caratteristica per generare i piani di navigazione di un robot in un museo archeologico: il robot decide quali sale visitare e quale percorso scegliere immaginando il percorso mediante una simulazione 3D del museo stesso. Il robot quindi è in grado di utilizzare concetti ipotetici del tipo "se… allora". Una volta generato il piano di navigazione, il robot poi verifica la sua esecuzione confrontando le scene reali acquisite tramite la sua telecamera, con le scene immaginate durante la pianificazione. Un approccio simile è stato seguito da Bongard *et al.* (2006) e da Holland *et al.* (2007).

In Chella *et al.* (2003) lo spazio concettuale è stato ulteriormente esteso per includere una rappresentazione semplificata del sé del robot: il robot è in grado non solo di descrivere le azioni percepite all'esterno effettuate da altre persone, ma anche di descrivere le proprie azioni osservando sé stesso. Si veda a questo proposito il lavoro di Hart, Scassellati (2015). Ad esempio, il robot, osservando i movimenti della propria mano, è in grado di descrivere il fatto che esso stia colpendo un oggetto con il dito. In questo modo il robot è in grado di utilizzare concetti relativi a sé stesso, come: "io sto colpendo una palla."

Nella rappresentazione del sé mediante spazi concettuali, il robot, esaminando il proprio spazio concettuale, è in grado di effettuare delle deduzioni, ad esempio che non

conosce il significato di un concetto, perché nel suo spazio concettuale manca l'insieme corrispondente. Quindi può dire che non sa come è fatto un certo oggetto. Allo stesso modo, il robot è in grado di dire che non sa come fare una determinata azione perché nel suo spazio concettuale mancano le strutture corrispondenti. Quindi, l'esame degli spazi concettuali porta il robot a utilizzare concetti di negazione come "non so come è", "non so come fare."

Infine, il robot può rappresentare nel suo spazio concettuale sé stesso che sta compiendo l'azione (Chella *et al.*, 2008). La figura 2 (a destra) mostra lo spazio concettuale del robot che percepisce un ostacolo (nella figura in basso): sono presenti il punto che rappresenta il robot e il punto che rappresenta l'ostacolo. La stessa figura a destra in alto mostra lo spazio concettuale del robot in cui il punto rappresenta tutta la situazione descritta precedentemente: il robot che osserva sé stesso.

Una ulteriore generalizzazione consiste nel fatto che il robot non solo può riflettere sull'azione che sta svolgendo, ma su tutte le sue azioni passate. Il robot può quindi utilizzare concetti legati alla sua storia passata, come ad esempio al fatto che prima non sapeva come fare un'azione ma adesso ha imparato a farla, o che prima poteva compiere una determinata azione ma adesso, a causa di un malfunzionamento dei motori o di sue limitazioni fisiche, non è in grado di compierla. Il robot potrà quindi dire "non posso farlo" perché inferisce dalle sue esperienze passate che non è in grado di farla.

Lo spazio concettuale relativo al robot, oltre alle caratteristiche fisiche del robot (la sua forma, il suo colore, la sua posizione, ecc.) rappresenta allora anche la memoria del robot stesso, quindi rappresenta la sua esperienza attuale ma anche le sue esperienze passate, ossia tutto quello che il robot ha percepito finora. Un altro esempio di robot

in grado di riflettere su se stesso e sulla propria conoscenza, ma basato sui formalismi della logica e non sugli spazi concettuali, è descritto da Bringsjord (2015).

Si noti che la rappresentazione del sé del robot cresce insieme all'accumulo di esperienze del robot: un robot appena uscito dalla fabbrica avrà un piccolo insieme di esperienze, e la sua percezione sarà immediata. Man mano che il robot accumulerà esperienze, avrà una rappresentazione di sé sempre più sofisticata. Due robot identici, potranno quindi avere una percezione immediata simile, perché stanno percependo gli stessi oggetti. Ma i due robot, avendo avuto esperienze passate diverse, avranno una percezione di sé differente e potranno quindi interpretare in maniera diversa la stessa scena. Ad esempio, un robot potrà ricordare una scena simile già vista, mentre per l'altro robot sarà una scena nuova.

4. Conclusioni

In questo capitolo abbiamo analizzato e messo a confronti tre diversi livelli di rappresentazione dei concetti: il livello simbolico, il livello neurale ed il livello intermedio degli spazi concettuali. Al livello simbolico, i concetti sono rappresentati attraverso definizioni, quindi mediante relazioni con altri concetti. Al livello neurale un concetto è rappresentato mediante l'attivazione di unità neurali artificiali. Tra i due livelli abbiamo consideriamo un livello intermedio di rappresentazione dei concetti basato sugli spazi concettuali.

Nella storia dell'intelligenza artificiale e delle scienze cognitive, tutti e tre gli approcci sono stati utilizzati per una varietà di applicazioni su sistemi esperti, programmi per riconoscimento oggetti e persone, e in robotica. Ma di

recente, soprattutto per il successo delle reti neurali profonde (*deep neural networks,* LeCun *et al.* 2015), cioè reti con numerosi strati di unità nascoste, come le reti convoluzionali, che vengono addestrate con enormi quantità di esempi di addestramento (per esempio, basi di dati presi dal web), vi è stato un significativo incremento dell'uso metodi di rappresentazioni neurali. Inoltre, questi approcci neurali sono adatti per l'apprendimento di concetti negli agenti robotici.

Il capitolo inoltre ha mostrato come l'uso della robotica dello sviluppo, collegata al metodo delle reti neurali per l'apprendimento delle parole, consenta al robot di acquisire concetti in maniera molto simile ai bambini, interagendo con un istruttore e con il mondo esterno. In particolare, questi esperimenti con robot dimostrano l'importanza di un approccio *emobodied* e *situated* (Cangelosi, Schlesinger 2015), alla simulazione della mente. Cioè al valore di contestualizzare l'apprendimento dell'agente cognitivo nel suo ambiente fisico e sociale (*situated*), e sfruttando l'interazione tra il proprio sistema sensomotorio e corporeo (*embodied*) e l'ambiente fisico, come dimostrato nel caso del ruolo della postura nell'acquisizione delle prime parole in bambini e robot.

Infine, il capitolo ha presentato esempi di utilizzo degli spazi concettuali per la robotica, dalla descrizione di scene alla descrizione di azioni, introducendo infine anche aspetti legati all'introspezione del robot. Gli spazi concettuali sono una sorta di lingua franca (Lieto *et al.*, 2017) che permette di unificare e integrare in un terreno comune le rappresentazioni simboliche e gli approcci subsimbolici e di superare alcuni noti problemi di tali rappresentazioni. In particolare, l'interpretazione degli spazi concettuali è molto più trasparente rispetto alle rappresentazioni basate sul-

le reti neurali, la cui interpretazione basata sull'analisi dei pesi delle connessioni neurali è generalmente controversa.

Inoltre, gli spazi concettuali offrono una ricchezza espressiva maggiore rispetto alle rappresentazioni simboliche. Ai pensi ad esempio alla rappresentazione dei concetti mediante prototipi o similarità, la cui descrizione a livello simbolico è tipicamente onerosa.

In conclusione, i sistemi intelligenti interattivi come i robot umanoidi o le macchine intelligenti richiedono la capacità da parte delle macchine di percepire il mondo esterno e di rappresentarlo mediante concetti. Non esiste un livello privilegiato di rappresentazione dei concetti ma sono necessari diversi livelli di rappresentazione quali il livello simbolico, il livello subsimbolico o connessionista ed il livello concettuale.

Riconoscimenti

Questo lavoro è stato in parte finanziato dal progetto della Unione Europea H2020 chiamati "DCOMM" e "APRIL" e dai progetti finanziati dalla US Air Force Office of Scientific Research intitolati THRIVE e "Self-Consciousness and Theory of Mind for a Robot Developing Trust Relationships."

Riferimenti bibliografici

BONGARD J., ZYKOV V., LIPSON H., (2006) *Resilient machines through continuous self-modeling*, «Science», 314:1118–1121.

BRINGSJORD S. (2015). *Real Robots That Pass Human*

Tests of Self-Consciousness. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2015*:498–504.

CANGELOSI A., GRECO A., HARNAD S., (2000) *From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories*, «Connection Science», 12(2):143–162.

CANGELOSI A., SCHLESINGER M., (2015) *Developmental Robotics: From Babies to Robots*, MIT Press, Cambridge, MA.

CHELLA A., DINDO H., INFANTINO I., (2006) *A cognitive framework for imitation learning,* «Robotics and Autonomous Systems», 54(5):403–408.

CHELLA A., FRIXIONE M., GAGLIO S., (1997) *A cognitive architecture for artificial vision,* «Artificial Intelligence», 89(1–2):73–111.

CHELLA A., FRIXIONE M., GAGLIO S., (2000) *Understanding Dynamic Scenes, «*Artificial Intelligence», 123:89–132.

CHELLA A., FRIXIONE M., GAGLIO S., (2003) *Anchoring symbols to conceptual spaces: the case of dynamic scenarios*, «Robotics and Autonomous Systems», 43(2-3):175–188.

CHELLA A., FRIXIONE M., GAGLIO S., (2008) *A cognitive architecture for robot self-consciousness,* «Artificial Intelligence in Medicine», 44(2):147–154.

CHELLA A., MACALUSO I., (2009) *The perception loop in CiceRobot, a museum guide robot* «Neurocomputing», 72(4–6):760–766.

CORADESCHI S., SAFFIOTTI A., (2003) *An Introduction to the Anchoring Problem*, «Robotics and Autonomous Systems», 43:85–96.

DE LA CRUZ V.M., DI NUOVO A., DI NUOVO S., CANGELOSI A., (2014) *Making fingers and words count*

in a cognitive robot, «Frontiers in Behavioral Neuroscience», 8:13.

GÄRDENFORS P., (2000) *Conceptual Spaces: The Geometry of Thought*, MIT Press, Bradford Books, Cambridge, MA.

GÄRDENFORS P., (2014) *The Geometry of Meaning: Semantics Based on Conceptual S*paces, MIT Press, Bradford Books, Cambridge, MA.

GRUSH R., (2004) *The Emulator Theory of Representation: Motor Control, Imagery and Perception,* «Behavioral and Brain Sciences», 27:377–442.

HARNAD S., HANSON S. J., LUBIN J., (1991), *Categorical Perception and the Evolution of Supervised Learning in Neural Nets*, in Powers D.W., Reeker L., (eds.): *Working Papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*, AAAI Press, Menlo Park, CA, pp. 65–74.

HARNAD S., HANSON S. J., LUBIN J., (1995), *Learned Categorical Perception in Neural Nets: Implications for Symbol Grounding,* in: Honavar V., Uhr L., (eds.): *Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration,* Academic Press, pp. 191–206.

HARNAD S., (1990) *The Symbol Grounding Problem*, «Physica D», 42:335–346.

HART J.W., SCASSELLATI B., (2015) *Robotic Self-Modeling*, in J. Pitt (eds.): *The Computer After Me*, Imperial College Press, London, pp. 207–218.

HEBB D.O., (1949). *The Organization of Behavior: A Neuropsychological Theory.* John Wiley & Sons, New York.

HESSLOW G., (2002) *Conscious thought as simulation of behaviour and perception*, «Trends in Cognitive Sci-

ences», 6:242–247.

HOLLAND O., KNIGHT R., NEWCOMBE R., (2007) *A robot-based approach to machine consciousness,* in Chella A., Manzotti R., (eds.): *Artificial Consciousness*, Imprint Academic, Exeter, UK, pp. 156–173.

KOHONEN, T., (2001), *Self-Organizing Maps*, Springer-Verlag, Berlin, Heidelberg, terza edizione.

LECUN Y., BENGIO Y., HINTON G., (2015) *Deep learning*, «Nature», 521:436–444.

LIETO A., CHELLA A., FRIXIONE M., (2017) *Conceptual Spaces for Cognitive Architectures: A lingua franca for different levels of representation*, «Biologically Inspired Cognitive Architectures», 19:1–9.

MCCLELLAND, J.L., RUMELHART D.E. (1981) *An Interactive Activation Model of Context Effects in Letter Perception: I. An Account of Basic Findings,* «Psychological Review», 88(5):375–407.

MORSE A.F., BENITEZ V.L., BELPAEME T., CANGELOSI A., SMITH L.B., (2015) *Posture Affects How Robots and Infants Map Words to Objects*, «PLoS ONE», 10(3):e0116012.

MORSE A.F., DEGREEFF J., BELPAEME T., CANGELOSI A., (2010) *Epigenetic Robotics Architecture (ERA)*, «IEEE Transactions on Autonomous Mental Development», 2(4):325–339.

NEWELL A., SIMON, H.A., (1972) *Human Problem Solving*, Prentice-Hall, Englewood Cliffs, New Jersey.

PARISI D., (1989) *Intervista sulle reti neurali. Cervello e macchine intelligenti,* Il Mulino, Bologna.

PLUNKETT K., SINHA C., (2011) *Connectionism and Developmental Theory*, «British Journal of Developmental Psychology», 10(3):209–254.

PLUNKETT K., SINHA, C., MØLLER, M.F., STRANDSBY, O.,

(1992) *Symbol Grounding or the Emergence of Symbols? Vocabulary Growth in Children and a Connectionist Net,* «Connection Science», 4(3-4):293–312.

ROSCH E., (1975) *Cognitive representations of semantic categories*, «Journal of Experimental Psychology: General», 104(3):192-233.

RUCINSKI M., CANGELOSI A., BELPAEME T., (2012) *Robotic model of the contribution of gesture to learning to count,* in: *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics,* pp. 1–6.

RUMELHART D.E., MCCLELLAND J.L. E THE PDP RESEARCH GROUP, (1986) *Parallel Distributed Processing, Explorations in the Microstructure of Cognition,* MIT Press, Bradford Books, Cambridge, MA.

SAMUELSON L.K., SMITH L.B., PERRY L.K., SPENCER J.P., (2011) *Grounding Word Learning in Space,* «PLoS ONE», 6(12): e28095.

SCHAAL S., (1999) *Is Imitation Learning the Route to Humanoid Robots?* «Trends in Cognitive Sciences» 3:233–242.

SCHAAL S., IJSPEERT A.J., BILLARD A., (2003) *Computational approaches to motor learning by imitation,* «Philosophical Transactions of The Royal Society: Biological Sciences», 358:537–547.

SOWA J., (1992) *Semantic Networks*, in Shapiro S., (eds.): *Encyclopedia of Artificial Intelligence,2nd edn.* J. Wiley & Sons, New York..

STRAMANDINOLI F., MAROCCO D., CANGELOSI A., (2017) *Making sense of words: a robotic model for language abstraction,* «Autonomous Robots», 41:367–383.

TIJSSELING A., HARNAD S., (1997), *Warping Similarity Space in Category Learning by BackProp Nets,* in: Ramscar M., Hahn U., Cambouropolos E., Pain H.,

(eds.): *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*, Department of Artificial Intelligence, Edinburgh, Scotland, pp. 263–269.

TWOMEY K.E., MORSE A.F., CANGELOSI A., HORST J.S., (2016) *Children's referent selection and word learning,* «Interaction Studies», 17(1):101–127.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130664
pp. 57-76 (marzo 2018)

Cleptomania

Genesi ed evoluzione di una categoria criminologica

di ROMOLO GIOVANNI CAPUANO[1]

1. Lo strano caso della signora Ella Castle

Il 14 novembre 1896, il «British Medical Journal» segnala ai suoi lettori un caso bizzarro. Un ricco commerciante di San Francisco si trova in vacanza in Europa con moglie e figlio. La coppia è la tipica rappresentante della ricca borghesia americana e occupa una posizione prominente in società. Al termine della vacanza, poco prima di far ritorno negli Stati Uniti, si scopre che nella loro stanza d'albergo, a Londra, è accumulata una gran quantità di merce rubata, di cui entrambi vengono incolpati. Il 5 ottobre 1896, la coppia viene tratta in arresto. Ben presto, le accuse contro il marito decadono e la donna, la signora Ella Castle, rimane l'unica colpevole tanto da essere condannata a tre mesi di prigione senza lavori forzati. La condanna, tuttavia, non viene eseguita in quanto la difesa fornisce un'importante spiegazione del comportamento della Castle: la donna è malata e non è, dunque, responsabile delle proprie azioni. Si dice che soffra di isteria e di "nervosismo" e che i suoi problemi risalgano all'infanzia. Un medico inglese la definisce "di temperamento e disposizione fortemente nervosi" (Abelson, 1989, p. 127). Soprat-

[1] Ricercatore indipendente. E–mail: romolo.capuano@gmail.com

tutto, il suo avvocato, Sir Edward Clarke, sostiene che soffre di "cleptomania", una condizione scoperta da poco che sembra colpire soprattutto le donne benestanti sorprese a rubare nei grandi magazzini. «Perché altrimenti una donna abbiente e di buona famiglia, con un marito pronto a soddisfare ogni suo bisogno, avrebbe dovuto rubare quattro oggetti di poco valore?» è la chiosa finale dell'avvocato. Tornata in patria, la donna è visitata da molti medici i quali concludono tutti che la donna soffre di turbe psichiche, classificabili come manie. Il processo di medicalizzazione del comportamento deviante della donna è portato a compimento. Il furto è solo il sintomo di una malattia. Non ha nulla a che fare con una volontà criminale. I veri delinquenti sono altra cosa. Perfino Arthur Conan Doyle, l'autore di Sherlock Holmes, interviene, con una lettera, a difesa della donna (Abelson, 1989, p. 134).

Il caso di Ella Castle non fu l'unico. Sul finire del XIX secolo, tante donne si ritrovarono a essere etichettate come "cleptomani". Ma che cosa era la cleptomania? E chi l'aveva scoperta?

2. Genesi di una categoria criminologica

La cleptomania fu identificata nel 1816 dal medico svizzero, Andre Matthey, il quale, in realtà, la battezzò con il nome di "clopemania". Secondo Matthey, la clopemania era una forma di monomania caratterizzata dall'impulso irresistibile a rubare. Più precisamente: «una mania unica caratterizzata dalla tendenza a rubare senza un motivo e senza necessità. La tendenza a rubare è permanente e non è associata ad alienazione mentale. La ragione conserva il suo dominio, si oppone a questa coazione segreta, ma la tendenza ha la meglio e la volontà ne è soggiogata» (cit. in

Fullerton, Punj, 2003, p. 202). All'epoca, il termine "monomania" era un termine ombrello che descriveva genericamente una manifestazione ossessiva per un oggetto. Coniando il termine "clopemania", Matthey restrinse il significato della mania agli oggetti rubati.

Il concetto trovò una decisiva elaborazione in seguito al lavoro del medico francese Charles Chrétien Henri Marc, il quale, nel 1840, coniò il termine "cleptomania" per designare «un impulso conscio a rubare che si verifica in un individuo in cui non vi è ordinariamente un disturbo della coscienza. L'individuo si oppone a questo impulso, ma la sua natura è irresistibile» (cit. in Fullerton, Punj, 2003, p. 202). Gli oggetti rubati sono spesso di poco valore, l'individuo non ha la necessità materiale di commettere il furto e la mania colpisce soprattutto persone appartenenti alle classi superiori, in particolare le donne. Per Marc, affinché si possa parlare di cleptomania, è necessario tener conto dell'esistenza di determinanti biologiche, come secrezioni ed escrezioni, ma anche di fattori sociali come il valore dell'oggetto rubato e la classe sociale di provenienza del ladro (Segrave, 2001, p. 20). La diagnosi, dunque, mescola indifferentemente elementi biologici e sociali, assunti entrambi come "dati" naturali.

La definizione di Marc fu resa prestigiosa dal suo maestro, il celebre psichiatra Jean–Étienne Dominique Esquirol, il quale, in un testo scritto con l'allievo, definì la cleptomania «una monomania consistente in una lesione della volontà di resistere» (Marc, Esquirol, 1838). Altri autori produssero definizioni più o meno simili: Crosby e Lunier nel 1879, Gross nel 1907, Friedemann nel 1930 (Fullerton, Punj, 2003, p. 202). Venti anni dopo Marc, Bénédict Augustin Morel introdusse il termine "degenerazione" come fattore esplicativo della cleptomania, mentre Janet e Ray-

mond assimilarono questa a una forma di comportamento isterico (O'Brien, 1983, p. 71).

Ma è soprattutto alla fine del XIX secolo che la categoria di "cleptomania" si afferma. Come sintetizza Patricia O'Brien:

> Dopo il 1880, la cleptomania è riconosciuta quasi universalmente come "un impulso morboso a rubare oggetti perfettamente inutili o che potrebbero agevolmente essere acquistati". Alla determinazione della diagnosi contribuiscono pesantemente sia manifestazioni di monomania sia manifestazioni di isteria. Si tratta di categorie onnicomprensive, sotto cui ricade una grande varietà di comportamenti che confluiscono in un repertorio generale che sottolinea la perdita della ragione. La diagnosi si basa sull'assunto che è irrazionale rubare oggetti di cui non si ha bisogno. Lo specialista forense conclude che chi si comporta così non è legalmente responsabile dei propri atti. Una cospicua letteratura psichiatrica sostiene il principio della irresponsabilità legale del cleptomane nella commissione del reato e ha un impatto determinante sul destino giudiziario delle borghesi sorprese a rubare nei negozi: quasi tutte […] sono assolte dai tribunali (O'Brien, 1983, p. 67).

Il fatto straordinario è che gli autori che si occuperanno di cleptomania in seguito al lavoro di Matthey e Marc potranno non essere concordi sull'eziologia del fenomeno e su come curarlo, ma non sulla sua definizione. Ancora oggi, le definizioni che si danno della cleptomania sono fondamentalmente concordi con gli esiti delle prime osservazioni psichiatriche sull'argomento (Fullerton, Punj, 2003, p. 202). Basta leggere la definizione che del fenomeno dà il DSM–5, la versione più recente del Manuale diagnostico e statistico dei disturbi mentali, la bibbia degli psichiatri di tutto il mondo. In questo testo, la cleptomania è fatta rientrare tra i disturbi dirompenti, da discontrollo degli impulsi

e della condotta ed è caratterizzata da cinque tratti (Biondi, 2014):

a) ricorrente incapacità di resistere all'impulso di rubare oggetti di cui non c'è bisogno per l'uso personale o per il loro valore economico;
b) crescente sensazione di tensione immediatamente prima di commettere il furto;
c) piacere, gratificazione o sollievo nel momento in cui il furto viene commesso;
d) il furto non viene compiuto per esprimere rabbia o vendetta, né in conseguenza di un delirio o un'allucinazione;
e) il furto non è meglio spiegato dal disturbo della condotta, da un episodio maniacale o dal disturbo antisociale di personalità.

Insomma, da Matthey e Marc a oggi non molto sembra cambiato nel modo di definire la cleptomania. Ciò sembra testimoniare a favore di una persistenza nel tempo dei sintomi e della diagnosi; una costanza che pare autorizzare una interpretazione "naturalistica" di questa condizione. In altre parole, se tanti medici e psichiatri, sin dall'Ottocento, discutono di cleptomania negli stessi termini è perché "evidentemente" a essa corrisponde una patologia accertata e indubitabile. Ma le cose stanno davvero così?

3. Teorie eziologiche della cleptomania

Se la definizione del termine "cleptomania" è rimasta stabile nel tempo, più variegato, come si è detto, è il quadro eziologico della condizione che si è prestato, nel corso del tempo, a interpretazioni anche molto diverse secondo la

scuola di pensiero dell'interprete. Se Marc ed Esquirol parlarono di "lesione della volontà" e di isteria, altri autori indicarono nell'epilessia, nella depressione, nell'alcolismo, nei difetti cerebrali, nella menopausa, nell'imbecillita o nella "atmosfera dei grandi magazzini" la causa principale del fenomeno. Nel 1839, Isaac Ray associò la cleptomania all'idiotismo e a possibili traumi cerebrali. Nel 1874, Henry Maudsley, uno dei fautori della teoria della degenerazione, descrisse la cleptomania come un tipo di "imbecillità morale" associata all'idiotismo che rendeva chi ne era affetto non responsabile delle proprie azioni (Whitlock, 2005).

Particolarmente creative sono le spiegazioni fornite dagli psicoanalisti. Secondo Wilhem Stekel (1911), un allievo di Freud, la cleptomania è la risposta a una forte frustrazione sessuale, mentre, negli uomini, segnala tendenze omosessuali. Per Alfred Adler, la cleptomania è la risposta a un sentimento di inferiorità fisica e sociale che provoca una intensa nevrosi. Per Fritz Wettels, i cleptomani sono soggetti sessualmente sottosviluppati con poche relazioni sessuali. Altri autori come Franz Alexander, Otto Fenichel e Sandor Rado chiamarono in causa concetti classici della psicanalisi come il complesso di castrazione, il complesso di Edipo e l'invidia del pene (Fullerton, Punj, 2003, p. 206). Tali spiegazioni vennero progressivamente abbandonate in ragione della progressiva perdita di prestigio della psicanalisi stessa.

La seconda edizione del Manuale diagnostico e statistico dei disturbi mentali (1968) epurò la nozione di cleptomania dalle sue pagine, per poi reintrodurla nella terza edizione e confermarla, come abbiamo visto, nella quinta. La fortuna della categoria ha poi oscillato nel tempo, subendo alti e bassi, ma rimanendo comunque come categoria medico–diagnostica.

Oggi, si è sostanzialmente concordi, in ambito medico e psichiatrico, nel sostenere che la diagnosi di cleptomania è tendenzialmente residuale ed è forse applicabile solo al 5% di tutti i taccheggiatori tratti in arresto. Permane, tuttavia, un certo scetticismo. Molti psichiatri esprimono seri dubbi sulla esistenza di una condizione psicologica che possa essere legittimamente e scientificamente definita "cleptomania" e sulla possibilità di una diagnosi che possa essere "oggettiva". Psichiatri come Stanton Samenow, ad esempio, affermano pubblicamente che «in 40 anni di attività con svariati trasgressori, compresi uomini e donne che rubano ripetutamente, non mi sono mai imbattuto in un vero caso di cleptomania» (Samenow, 2011, p. 1). Gli stessi dubbi sono condivisi da autorità legali, criminologi ed esperti di varia provenienza. In alcuni casi, si sospetta che la diagnosi sia "costruita" in modo da conformarsi allo stereotipo della patologia per evitare conseguenze penali (Krasnovsky, Lane, 1998, p. 223). Nella vita quotidiana e in molti articoli giornalistici, infine, il termine è spesso usato come equivalente improprio del termine "taccheggio", generando confusione semantica negli osservatori.

4. Cleptomania o taccheggio?

La cleptomania è qualcosa di diverso dal taccheggio (*shoplifting*, in inglese) che pure cominciò a diffondersi nella stessa epoca in cui Marc inventò la sua patologia. Oggi il taccheggio appare come uno dei reati più diffusi nel mondo occidentale, anche se è al contempo uno dei più sottovalutati (Clarke, Petrossian, 2013; Cupchik, Atcheson, 1983). Le ricerche dimostrano che chi si dedica al taccheggio non è necessariamente un cleptomane e, soprattutto, non corrisponde necessariamente allo stereotipo

della donna alto–borghese. Per quanto sia difficile avere dati precisi sul fenomeno, in quanto la propensione a denunciare delle vittime (negozi, grandi magazzini ecc.) è molto bassa, si sa, dai pochi studi empirici effettuati che, ad esempio, in Italia, solo una quota modesta di clienti si dedica al taccheggio (meno del 2%) e che tra questi figurano non solo donne e adolescenti (come vuole un altro stereotipo) ma anche ultracinquantenni (Barbagli, Colombo, Savona, 2003, p. 145).

Negli Stati Uniti, la provenienza sociale degli *shoplifters* non è dissimile da quella della clientela nel complesso. Tra essi è possibile trovare giovani e vecchi, donne e uomini, individui isolati e gruppi organizzati, persone rispettabili e individui "poco raccomandabili". Nella stragrande maggioranza dei casi, le motivazioni non sono riconducibili a impulsi irrefrenabili, ma a povertà, disoccupazione, opportunità, bassa percezione di rischio, vendetta, rabbia e altri fattori non patologici. Una famosa ricerca compiuta negli Stati Uniti negli anni Quaranta del XX secolo mise in evidenza che esistevano due categorie di taccheggiatori: i *boosters* e gli *snitches*. I *boosters* (il 10% dei taccheggiatori) erano ladri professionisti, appartenenti a sottogruppi criminali, coinvolti in altre attività illegali. Gli *snitches* (il 90% dei taccheggiatori), al contrario, erano cittadini "rispettabili" che non rubavano per motivi utilitaristici né perché affetti da cleptomania e che, una volta scoperti, difficilmente tornavano a rubare (Cameron, 1964).

Da alcune ricerche più recenti emerge chiaramente che non è corretto opporre ai furti irrazionali dei cleptomani quelli razionali e utilitaristici del taccheggiatore. Ad esempio, lo studio di alcuni minorenni taccheggiatori ha permesso di riscontrare in alcuni di essi, la presenza di disturbi di personalità, insufficienze intellettive, nevrosi, disturbi dell'umore, immaturità decisionale ed emotiva, di-

sadattamento, insuccesso scolastico, emarginazione, a volte appartenenza a sottoculture (Birkhoff, Pieri, Tavani, 2007, p. 99). Alcune casalinghe, invece, sembrerebbero rubare

> per una forma di risentimento nei confronti della propria condizione di vita, specie al raggiungimento dell'età media, quando possono presentare una tendenza alla depressione e una minore tolleranza alla frustrazione, forse scaturite dal venir meno delle loro funzioni di madri o per un minor interesse mostrato loro dai mariti (Birkhoff, Pieri, Tavani, 2007, p. 99).

In altri casi, interverrebbe una delinquenza "ludica", senza fini utilitaristici, scaturita dalla ricerca di una particolare tensione emotiva, connessa al brivido del rischio, o per acquisire prestigio nel gruppo, ovvero per noia o mancanza di impegni.

Il taccheggio, infine, costituisce una quota trascurabile dell'ammanco subito dai negozianti e non supera spesso l'uno per cento del totale del valore dei beni acquistati dai clienti in maniera lecita (uno dei motivi per cui le denunce sono così poche) (Barbagli, Colombo, Savona, 2003, p. 145).

5. La prospettiva costruzionista in criminologia

Secondo la prospettiva costruzionista, i fenomeni sociali non esistono "oggettivamente", ma perché qualcuno li definisce tali e li "costruisce" in base a interessi, motivazioni, risorse, convinzioni e con conseguenze che variano da situazione a situazione. Per dirla con il sociologo Stuart Henry,

> dal punto di vista del costruzionismo sociale, il crimine è una classificazione del comportamento definita da individui dotati del potere e dell'autorità di creare leggi che identificano un certo comportamento come trasgressivo e ne sottopongono a sanzione l'autore. Nelle società occidentali, i legislatori e i tribunali, coadiuvati dalle agenzie di stato, hanno il potere e l'autorità di definire il crimine e infliggere la punizione. La definizione di determinate forme di comportamento come crimini riflette sia i loro valori e interessi sia le norme e i valori collettivi della società, o almeno delle categorie più intraprendenti della società (Henry, 2009, pp. 1–2).

Secondo questa prospettiva,

> un problema sociale esiste quando: (1) un gruppo o categoria sociale considera anomala una data condizione o fenomeno; (2) il gruppo o categoria sociale esprime preoccupazione per l'esistenza della condizione o fenomeno; (3) il gruppo o categoria sociale invita a prendere provvedimenti o prende provvedimenti per correggere a condizione o fenomeno (Goode, Ben–Yehuda, 2009, p. 151).

Detto in altro modo, una determinata condizione diviene un problema sociale quando un gruppo di individui intraprende azioni di rivendicazione (*claims–making*) nei confronti di quella condizione, intendendo per "rivendicazione" «la richiesta espressa da un gruppo a un altro di fare qualcosa a proposito di una condizione putativa» (Spector, Kitsuse, 1977, p. 78). L'aggettivo "putativa" sottolinea il fatto che è la percezione della condizione a essere rilevante indipendentemente dal fatto che i meriti della rivendicazione siano veri o falsi.

L'approccio costruttivista consente di interpretare le categorie delle scienze sociali non come un "dato" definito una volta per tutte, ma come un artefatto, edificato in un certo tempo e luogo, di cui è importante intendere le modalità e i fini della costruzione. Una implicazione è che gli

artefatti, così come sono edificati, possono essere decostruiti per rilevarne la storicità, le criticità e le modalità di costruzione. Ciò consente di assumere un atteggiamento critico nei confronti dei fenomeni sociali: un approccio che non accetta supinamente la "datità" dei fenomeni sociali, ma ne mette in discussione la genesi e le funzioni.

Adottando la prospettiva costruzionista, categorie diagnostiche comunemente applicate a fatti devianti possono essere rappresentate nei loro aspetti storici, criminologici, psichiatrici e morali, ricostruendo i dibattiti intellettuali e disciplinari che, dalla data della loro origine, si sono sviluppati intorno a essi e le oscillazioni, talvolta profonde, di natura epistemologica, culturale e sociale che tali categorie subiscono nel corso del tempo. In sintesi, secondo la prospettiva costruzionista, la realtà sociale è costruita sulla base delle istanze delle classi dominanti e il lavoro dello scienziato sociale è di "scoprire" tali costruzioni (spesso credute naturali e immutabili dalle classi subalterne).

6. Costruzionismo e cleptomania

È possibile applicare la prospettiva costruzionista alla categoria di "cleptomania"? È possibile, in altre parole, individuare, in questo caso, come sostengono Goode e Ben–Yehuda, un gruppo o categoria sociale che considera anomala una data condizione o fenomeno, esprime preoccupazione per l'esistenza della condizione o fenomeno e invita a prendere provvedimenti o prende provvedimenti per correggere la condizione o fenomeno? Per rispondere a queste domande è necessario inquadrare la genesi della categoria nel suo contesto storico e sociale.

La nascita della cleptomania avviene in un determinato contesto storico e sociale: gli Stati Uniti e l'Europa della

seconda metà del XIX secolo; presuppone un determinato sistema economico: il capitalismo, orientato non più solo alla produzione, ma anche al consumo di massa, seppure ancora incipiente; consumo che trova il suo teatro di riferimento in una invenzione dell'epoca: il grande magazzino (*department store)*; coincide con l'importanza sempre più grande attribuita alla scienza e, in particolare, alla medicina forense; presuppone, infine, una precisa divisione di classe e una rigida divisione dei ruoli di genere che affida le attività di consumo alle donne.

È in questo contesto che si verifica una condizione anomala, che sembra sfuggire a ogni interpretazione razionale e alla quale una categoria sociale – quella dei medici – è chiamata a trovare una soluzione. La seconda metà del XIX secolo, in particolare gli anni successivi al 1880, assistono a un moltiplicarsi esponenziale di un nuovo comportamento criminale, la cui novità riguarda sia l'attore del crimine (la donna), sia la classe sociale di provenienza del criminale (la borghesia medio–alta), sia il tipo e il luogo del crimine (il furto nel grande magazzino), la cui diffusione esplode proprio in quegli anni tanto da indurre psichiatri come Alexandre Laccasagne a parlare di "fenomeno sociale" (O'Brien, 1983, pp. 65–66).

Incrociando le coordinate del contesto storico–sociale, componendole acrobaticamente con la condizione anomala che si verifica in quegli anni e che pone all'attenzione dell'opinione pubblica un enigma culturale senza precedenti, la nozione di cleptomania nasce allo scopo precipuo di "spiegare", in termini di malattia mentale, un comportamento – l'anomalia – che vede protagoniste donne borghesi che commettono furti nei grandi magazzini senza un motivo apparente. Con un colpo di bacchetta magica, la nozione di cleptomania permette di compiere una straordinaria *reductio ad unum*, in cui una serie di "astrusità" di

classe, di genere, sociali e culturali trovano improvvisamente una loro ragione. In altre parole, la nozione di cleptomania fu inventata per spiegare ciò che i contemporanei avvertivano come inspiegabile alla fine del XIX secolo e per sopire le ansie degli stessi relative ai profondi cambiamenti sociali, di genere e di consumo che si stavano verificando in quegli anni.

Innanzitutto, il furto perpetrato nei negozi dalle donne borghesi lasciava perplessi benpensanti e criminologi perché sembrava contraddire le idee tradizionali di "borghesia" e di "comportamento criminale". Come spiega Patricia O'Brien,

> il sapere convenzionale spiegava il furto soprattutto in termini di bisogni e deprivazioni o di sciagure morali associate a questi fattori. Come abbiamo osservato nell'opera di Marc e dei successivi specialisti forensi, l'assenza del bisogno era diventata la principale caratteristica che contraddistingueva il furto nei negozi perpetrato dalle donne borghesi. Si avvertiva la necessità di una nuova spiegazione: se uomini e donne della classe operaia rubavano per bisogno o per avidità, le donne borghesi rubavano perché erano malate (O'Brien, 1983, p. 73).

In secondo luogo, la categoria del cleptomane – o meglio della cleptomane – sfidava i criminologi anche su un altro terreno: essa metteva in dubbio l'assunto di tipo sociologico secondo cui è l'ambiente a produrre il delinquente. Le cleptomani provenivano da classi sociali rispettabili e altolocate che non inducevano "naturalmente" al furto. Di qui la necessità di una spiegazione alternativa in chiave medica e personale, che consentisse di trovare una spiegazione anche a questa evidente contraddizione di uno dei luoghi comuni ritenuti più affidabili dagli esperti (O'Brien, 1983, p. 74).

Ancora, la cleptomane metteva in discussione l'assunto che le classi borghesi fossero ambienti sociali pacifici e ordinati. I furti di tante donne perbene lasciavano sospettare famiglie problematiche e alienate, rapporti coniugali in crisi, problemi psichici e relazionali. Tutto il contrario dell'elogio della borghesia che veniva fatto in quegli anni dal senso comune. La nozione di cleptomania consentiva di risolvere questo ulteriore enigma, spostando la responsabilità dell'anomalia dalla struttura di classe e di genere borghese a un disturbo della mente femminile. La prima era, così, fatta salva a spese della donna, o meglio, di alcune donne (O'Brien, 1983, p. 74).

Contemporaneamente, il concetto di cleptomania servì da critica medica al nascente consumismo, avvertito come un fattore che stimola impulsi criminali nei cittadini rispettabili e che seduce diabolicamente al nuovo imperativo dell'acquisto a ogni costo. Non a caso lo psichiatra Alexandre Laccasagne definiva i grandi magazzini come "posti pericolosi" (O'Brien, 1983, p. 73) che stimolano, confondono e affascinano il cliente, inducendolo a comportamenti antisociali:

> Come l'assenzio e il vermut stimolano l'appetito per il cibo, così i banconi stipati di merci accrescono la brama femminile del possesso. Perfino le donne con maggiore forza di volontà cedono, spendendo più di quanto avrebbero fatto per le proprie necessità se fossero state assennate. Ma chi è in grado di misurare la forza che attrae e domina le menti deboli o degenerate? (cit. in Dominguez, 2009).

Anche lo scrittore Émile Zola, nel suo celebre *Il paradiso delle signore* (1883), cita il caso delle ladre per mania che rubano «per una perversione del desiderio, una nuova nevrosi descritta da un alienista come effetto patologico della tentazione esercitata dai grandi magazzini» (Zola,

2017, p. 285). In questo senso, la nozione di cleptomania consentiva di delineare una patogenesi sociale ed esprimere un monito a questa legato: Attenti ai grandi magazzini che confondono e seducono!

Soprattutto, però, la nozione di cleptomania consentiva di ristabilire i rapporti di genere su binari tradizionali, soffocando il nuovo protagonismo e l'indipendenza dagli uomini che il consumo di merci nei grandi magazzini assegnava alle donne e lasciando palesare i pericoli medici e, dunque, "naturali" a cui poteva esporre tale indipendenza. La patologizzazione del corpo femminile venne formulata facendo ricorso soprattutto a fattori sessuali (problemi mestruali, gravidanze, menopause ecc.) ed ereditari (malattie mentali, suicidi in famiglia, malformazioni fisiche, alcolismo ecc.) di cui il furto divenne irrimediabilmente il sintomo. In alcuni casi, lo psichiatra credette di rinvenire nel gesto criminale della donna un significato sessuale simile all'orgasmo. La cleptomania, in altre parole, confermava il ruolo subordinato della donna, contribuendo, fra l'altro, alla femminilizzazione dell'atto deviante e all'idea che le donne fossero più interessate degli uomini agli "oggetti materiali" e che questi ultimi riuscissero a dominare i loro impulsi più delle donne.

Insomma, non è esagerato dire che la diagnosi di cleptomania fu edificata su una serie di assunti preesistenti di classe e di genere che influenzarono irrimediabilmente la teoria medica (Whitlock, 2005, pp. 435–436). L'efficacia della diagnosi di cleptomania testimonia, inoltre, il potente ruolo che la medicina ottocentesca aveva nell'orchestrare, plasmare e dare visibilità analitica alle differenze di genere, conferendo a esse la forza di un dato naturale, non discutibile.

Il successo della categoria di "cleptomania" fu dovuto indubbiamente al largo uso che, nella seconda metà del

XIX secolo, di essa si fece in ambito giudiziario come strategia difensiva per far assolvere donne borghesi accusate di crimini grossolani e altrimenti inspiegabili. Già all'epoca, però, alcune voci critiche si levarono a contestarne la validità scientifica. Tra queste possiamo ricordare quella dello scrittore Mark Twain il quale, in un pezzo intitolato *A New Crime*, osservava ironicamente che «al giorno d'oggi, inoltre, se un individuo di buona famiglia e di alta condizione sociale ruba qualcosa, viene definito cleptomane e mandato al manicomio» (Twain, 1875); oppure quella dell'anarchica Emma Goldman che, in un discorso pronunciato nel 1896 a Pittsburgh, denunciò la cleptomania come uno stratagemma delle classe abbienti per derubare i poveri (Shteir, 2011). Dal canto suo, il medico Sir John Charles Bucknill non ebbe esitazioni a definire la nuova diagnosi come un "triste miscuglio" di varie osservazioni costrette con la forza in una unica definizione (Bucknill, 1863). Bucknill notò anche che spesso le argomentazioni degli psichiatri si reggevano sempre sugli stessi vecchi casi, citati in continuazione a sostegno dell'esistenza del disturbo. Espresse, infine, preoccupazione per l'abuso giudiziario della categoria; preoccupazione condivisa anche da altri commentatori, convinti della necessità di punire chi si macchiava di reati banali, anche se commessi da esponenti della borghesia medio–alta (Fullerton, Punj, 2003, p. 203).

7. Conclusioni

Come rilevano Lenz e MagShamhráin:

> Le teorie e le pratiche mediche non sono mai quelle faccende meramente pragmatiche o scientifiche che sembrano. Sebbene

> gli esperti che le articolano e le propongono le considerino spesso come se fossero del tutto incontaminate dalle loro matrici culturali, politiche e sociali, esse sono profondamente radicate nel loro tempo e nella loro cultura (Lenz, MagShamhráin, 2012, p. 279).

Seguendo queste indicazioni, la categoria di cleptomania può essere letta come una reazione alla modernizzazione radicale della società che si verificò tra il XIX e il XX secolo, e come un meccanismo di controllo sociale concepito per patologizzare le donne che, grazie ai grandi magazzini, ebbero storicamente, per la prima volta, la possibilità di sfuggire al controllo maschile attraverso l'atto del consumo, ma anche attraverso il lavoro: molte donne, infatti, vennero impiegate nei grandi magazzini come commesse[2]. La critica alla nuova istituzione del consumo si tradusse, dunque, nella invenzione di un germe patogeno che faceva ammalare le donne, rendendole imprevedibili, instabili e propense al crimine.

La nozione di cleptomania, però, consentì anche di confermare gli assunti dell'epoca riguardanti i rapporti tra le classi sociali, le idee relative al ruolo criminogeno dell'ambiente e alla genesi del comportamento criminale. Permise, al tempo stesso, di scongiurare i timori relativi a una crisi sociale, culturale e sessuale della borghesia dell'epoca.

La nuova malattia incarnò, in un certo senso, i timori di un'epoca: timori relativi alla stabilità dell'ordine borghese e alla grande libertà che, improvvisamente, un gran numero di donne si ritrovarono ad avere grazie all'istituzione dei grandi magazzini; libertà che provocò agitazione in una società dominata da valori patriarcali. La cleptomania,

[2] Come è evidente, in maniera esemplare, dal romanzo di ZOLA, *Il paradiso delle signore* (1883).

come malattia, può, dunque, essere interpretata come metafora dei turbamenti della società moderna, i cui cambiamenti, troppo repentini, spaventavano le persone dell'epoca.

Riferimenti bibliografici

ABELSON E.S., (1989) *The Invention of Kleptomania*, «Signs», vol. 15, n. 1, pp. 123–143.

BARBAGLI M., COLOMBO A., SAVONA E., (2003) *Sociologia della devianza*, il Mulino, Bologna.

BIONDI M. (a cura di), (2014) *DSM–5. Manuale diagnostico e statistico dei disturbi mentali*, Raffaello Cortina Editore, Milano.

BIRKHOFF J. M., PIERI C., TAVANI M. (2007) *Il taccheggio: furto o che altro?*, «Rassegna italiana di criminologia», anno I, n. 2, pp. 95–110.

BUCKNILL J. C. (1863) *Kleptomania*, «Journal of Medical Science», vol. 8, n. 42.

CAMERON M.O., (1964) *The booster and the snitch: Department Store Shoplifting*, Free Press, New York

CLARKE R.V., PETROSSIAN G., (2013) *Shoplifting, Problem–Specific Guides Series n. 11, COPS* (Community Oriented Policing Services), Center for Problem–Oriented Policing, U.S. Department of Justice, Washington D.C.

CUPCHIK W., ATCHESON D. J., (1983) *Shoplifting: An Occasional Crime of the Moral Majority*, «Bulletin of the American Academy of Psychiatry and the Law», vol. 11, n. 4, pp. 343–354.

DOMINGUEZ D. V. (2009) *Manufacturing Kleptomania: The Social and Scientific Underpinnings of a Pathology*, «Madison Historical Review», vol. 6, art. 1.

FULLERTON R. A., PUNJ G. N., (2003) *Kleptomania: A Brief Intellectual History*, in Shaw E. H. (a cura di), *The Romance of Marketing History. Proceedings of the 11th Conference on Historical Analisys and Research in Marketing* (CHARM), Michigan State University, Michigan.

GOODE E., BEN–YEHUDA N., (2009) *Moral Panics. The Social Construction of Deviance*, Wiley–Blackwell, England.

HENRY S., (2009) *Social Construction of Crime*, in Miller J. (ed.), *21st Century Criminology: A Reference Handbook*, SAGE Publications, Thousand Oaks, CA.

Kleptomania: The Case of Mrs. Castle, «The British Medical Journal», vol. 2, n. 1872 (14 nov. 1896), pp. 1462–1463.

KRASNOVSKY T., LANE R. C. (1998) *Shoplifting: A Review of the Literature*, «Aggression and Violent Behavior», vol. 3, n. 3, pp. 219–235.

LENZ T., MAGSHAMHRÁIN R. (2012) *Inventing Diseases: Kleptomania, Agoraphobia and Resistance to Modernity*, «Society», vol. 49, pp. 279–283.

MARC C. C. H., (1840) *De la Folie Consideree dans ses Rapports avec les Questions Medico–Judicieres*, Balliere, Paris.

MARC C. C. H., Esquirol E. (1838) *Consultation sur un Cas de Suspicion de Folie, Chez une Femme Inculpee de Vol*, «Annales d'Hygiene Publique», vol 40, pp. 435–460.

O'BRIEN P., (1983) *The Kleptomania Diagnosis: Bourgeois Women and Theft in Late Nineteenth–Century France*, «Journal of Social History», vol. 17, n. 1, pp. 65–77.

SAMENOW S. E. (2011) *Kleptomania: A Reality or Psychiatric Invention?*, «Psychology Today», 4 marzo, dis-

ponibile presso:
https://www.psychologytoday.com/blog/inside-the-criminal-mind/201103/kleptomania-reality-or-psychiatric-invention.

SEGRAVE K., (2001) *Shoplifting. A Social History*, McFarland & Company, Jefferson, North Carolina.

SHTEIR R., (2011) *The Steal: A Cultural History of Shoplifting*, Penguin Books, New York.

SPECTOR M., KITSUSE J.I., (1977) *Constructing Social Problems*, Cummings, Menlo Park, CA.

STEKEL W., (1911) *The Sexual Root of Kleptomania*, «Journal of Criminal Law and Criminology», vol. 2, n. 2, pp. 239–246.

WHITLOCK T. C. (2005) *Crime, Gender and Consumer Culture in Nineteenth–Century England*, Ashgate, Burlington.

TWAIN M. (1875) *A New Crime*, in Idem, *Sketches New and Old*, American Publishing Company, Hartford (Connecticut) & Chicago (Illinois).

ZOLA É., (2017) *Il paradiso delle signore*, Mondadori, Milano.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130665
pp. 77-94 (marzo 2018)

Come le teorie cognitive possono aiutare l'intelligenza artificiale

di LUCIANO CELI[1]

1. Un breve preambolo storico

Che i robot abbiano segnato la storia del secolo passato – e non solo[2] – è un dato inconfutabile: nel 1912 i ricercatori John Hammond Jr. e Benjamin Miessner davano vita al primo prototipo di robot: una rudimentale (ai nostri occhi) scatola su ruote, che venne chiamata "cane elettrico" – forse perché fedelmente seguiva la luce proiettata verso i suoi sensori dai suoi creatori – che fu visto inizialmente come una curiosità scientifica[3]. Solo qualche anno dopo, il 25 gennaio 1921, andava in scena al teatro nazionale di Praga il dramma utopico-fantascientifico R.U.R. (sigla che sta per *Rossumovi univerzální roboti*, traducibile come *I robot universali di Rossum*), atto che sancisce l'ingresso definitivo della parola "robot" nel lessico mondiale, la cui definizione, in prima battuta, può essere appunto quella di un automa dotato di intelligenza artificiale (d'ora in poi AI).

[1] Dipartimento di Ingegneria Civile, Ambientale e Meccanica (DICAM), Università degli Studi di Trento e Istituto per i Processi Chimico–Fisici (IPCF) del Consiglio Nazionale delle Ricerche. E-mail: luciano.celi@unitn.it

[2] Per una rassegna storica si veda LOSANO (1991).

[3] Il "caso" è tra i più studiati ed è ampiamente descritto sia dagli autori, sia dagli studiosi che riportano fedelmente le descrizioni originali – si vedano in tal senso CORDESCHI (2002) e TAMBURRINI (2005).

La fantascienza ha sempre precorso, almeno in questo filone, ciò che, prima o dopo, sarebbe accaduto nella realtà, dipingendo di volta in volta scenari più o meno plausibili. Il "cane elettrico" di Hammond e Miessner non può non farci tornare alla mente le "pecore elettriche" di Philip K. Dick[4], osannato bestseller del genere e, nell'omologa trasposizione cinematografica[5], divenuto presto una pietra miliare, anche grazie alla sapiente regia di Ridley Scott.

Questa breve nota, apparentemente estranea a ciò che segue, è utile a mettere in mostra il potere che da sempre evoca questa sorta di "trasposizione divina", secondo la quale l'Uomo si sostituisce all'entità divina che si ipotizza lo abbia creato, per diventare a sua volta creatore *tout court* e, a propria immagine e somiglianza, realizzare esseri che sempre più gli si possano avvicinare nella struttura fisica o nel pensiero o, nel più complesso dei casi, in entrambe le sfere.

2. L'IA oggi

In molti settori l'utilizzo dell'IA, che fino a qualche decennio fa sembrava essere fantascienza, è divenuto realtà. Moltissimi passi avanti si sono fatti in settori molto specifici, come quello della *computer vision*: i medici possono essere coadiuvati da sistemi di IA – come Watson dell'IBM – per effettuare diagnosi, in alcuni casi più accurate di quanto il solo medico sarebbe riuscito a fare[6].

---

[4] Il titolo originale del libro, uscito nel 1968 negli Stati Uniti era infatti *Do Androids Dream of Electric Sheep?*, tradotto in italiano nel 1971 con un certamente meno evocativo *Il cacciatore di androidi*.

[5] *Blade runner*, 1982.

[6] Per una spiegazione sugli sviluppi di questo settore si rinvia FORNELL (2017) e ai video descrittivi ad esso collegati. Esistono inoltre delle statistiche

Da tempo è annunciata da più parti l'auto "che si guida da sola", capace quindi di riconoscere "ostacoli in movimento" come i pedoni. Se in questo specifico compito molte cose sembrano ancora dover essere messe a punto, in commercio si trovano già da qualche tempo auto che in dotazione hanno il *fatigue detection*, i cui studi, risalenti ormai a una decina d'anni fa[7], hanno permessi di realizzare un sistema di IA che effettua un monitoraggio costante del viso del guidatore e rileva i segni tipici della stanchezza.

Ovviamente molte altre sono le applicazioni in cui trova sempre più spazio l'IA, ma esulano da questo articolo.

2.1. *Il riconoscimento dei pattern*

L'apprendimento profondo (*deep learning*, d'ora in poi DL) nelle reti neurali è una tecnica ormai matura, che in anni recenti ha subito ulteriori accelerazioni. In particolare il *deep learning* permette ai modelli computazionali che sono composti da più livelli di elaborazione di apprendere le rappresentazioni di dati con più livelli di astrazione[8]. Questi metodi hanno migliorato considerevolmente lo stato dell'arte nel riconoscimento vocale, nel riconoscimento di oggetti visivi, nel rilevamento degli oggetti e molti altri domini. Basato sostanzialmente su un algoritmo di retroazione (*feedback*), questo agisce "all'indietro" (*backpropagation*) per indicare come la macchina debba cambiare i

che indicano come virtuosa l'interazione dell'uomo e dell'IA in questo compito specifico in relazione alle percentuali di successo nelle diagnosi. Percentuali che si abbassano di qualche punto percentuale se è solo l'uomo a compierle e di molti punti percentuali se si lascia l'intera attività decisionale all'IA.

[7] Cfr., per esempio, SENARATNE, HARDY, VANDERAA, HALGAMUGE (2007).

[8] Cfr. su questo argomento lo stato dell'arte di LECUN, BENGIO, HINTON (2015) e SCHMIDHUBER (2015). Nello specifico, per il caso del riconoscimento dei volti, ZHOU, CAO, YIN (2015).

propri parametri interni utilizzati per calcolare la rappresentazione per ognuno dei livelli di astrazione.

L'aspetto sconcertante – come hanno mostrato Anh Nguyen, Jason Yosinski e Jeff Clune in un articolo del 2015 – è che le reti neurali così performanti sono indotte a riconoscere immagini, create mediante un algoritmo evolutivo, prive di senso per un essere umano, identificandole come chitarre, pinguini reali, jackfruit, stelle marine e molto altro con una percentuale di confidenza, nella quasi totalità dei casi, di oltre il 99% – praticamente la certezza, come evidenziato in figura 1.

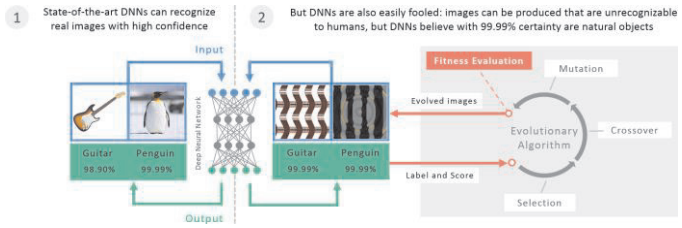

Figura 1. Anche se le reti neurali avanzate di ultima generazione possono riconoscere con sempre maggiore affidabilità le immagini naturali (pannello a sinistra), vengono anche facilmente ingannate nel dichiarare, con una confidenza che spesso è oltre il 99%, che le immagini non riconoscibili sono oggetti familiari (centro). Le immagini che ingannano le DNNs (Deep Neural Networks) sono prodotte da algoritmi evolutivi (pannello a destra) che ottimizzano le immagini per generare previsioni nella rete neurale con elevato grado di confidenza per ogni classe del set di dati su cui la rete stessa viene addestrata (qui, ImageNet).
Fonte: NGUYEN A., YOSINSKI J., CLUNE J. (2015)

L'avvento dei *big data* in ambito informatico è stato di grande aiuto per lo sviluppo generale del *machine learning*, di cui le reti neurali costituiscono un – se non il – settore di punta. Il meccanismo mediante il quale le reti apprendono è sostanzialmente costituito dalla iterazione di un certo numero di immagini, nel caso del riconoscimento

visivo, e, più alto è il numero di immagini sottoposte alla macchina, più "fine" sarà la sua capacità di discernimento e di categorizzazione nel modo corretto. Mentre questa capacità ha raggiunto livelli sorprendenti, anche su immagini complesse e contenenti molte informazioni, questa operazione sembra avere dei *bias* e dei totali disallineamenti con compiti semplici o immagini che, appunto, sono prive di significato per un essere umano.

Per quanto vi sia un'opinione scientifica prevalente in senso opposto – secondo cui, cioè, vi è ampio consenso sul successo degli algoritmi di DL[9] – resta da sottolineare che spesso lo scopo del riconoscimento e della categorizzazione in sistemi artificiali è – e soprattutto sarà sempre di più – di tipo applicativo: affidereste la vostra vita a un sistema che riconosce come "pinguino" un insieme arbitrario di linee e colori privo di significato all'occhio umano?

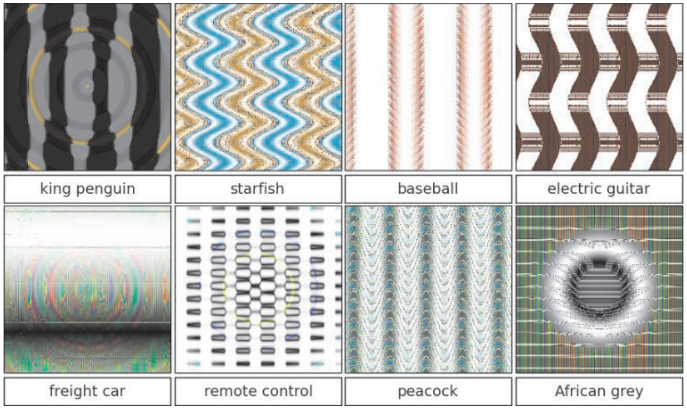

Figura 2. Altri esempi di immagini riconosciute da ImageNet come oggetti del mondo reale, senza senso per l'essere umano.
Fonte: NGUYEN A., YOSINSKI J., CLUNE J. (2015)

[9] VAN RULLEN (2017).

Un altro esempio ben documentato è costituito dal riconoscimento della grafia umana: numeri e lettere scritti da un essere umano, dopo il dovuto apprendimento, vengono riconosciuti con una confidenza molto alta e con elevato successo. Ma anche qui immagini senza senso, vengono riconosciute come numeri, come mostrato nella figura 3.

A onor del vero le immagini riconosciute dai sistemi di IA come significative sono sempre generate *ad hoc* e derivano da procedimenti di ottimizzazione molto sofisticati: sollevare dubbi però sulla reale affidabilità di questi sistemi è doveroso.

2.2. *Il funzionamento delle DNN*

I sostenitori della robustezza delle reti DL, ne mostrano la validità testandole sui "rumori di fondo", vale a dire, sulla capacità di riconoscere correttamente anche in presenza di elementi di disturbo contenuti all'interno dell'immagine[10], oppure su applicazioni reali, anche molto critiche[11]. Tutto questo però non sembra impedire il fatto che se da un lato abbiamo quella che potremmo definire una sovradeterminazione del riconoscimento ("*x* è riconosciuto come appartenente alla classe *X*, anche se realmente non lo è"), dall'altro è stato dimostrato anche il fenomeno contrario, quello della sottodeterminazione ("*x* non è riconosciuto come appartenente alla classe *X*, anche se realmente lo è"). Uno studio di Szegedy e colleghi[12] ha mostrato come questo fenomeno sia altrettanto frequente: sono state sufficienti delle piccole modifiche a una certa immagine – modifiche pressoché indistinguibili all'occhio umano, tanto

[10] Fawzi, Moosavi-Dezfooli, Frossard (2016).

[11] Lu, Sibai, Fabry, Forsyth (2017).

[12] In bibliografia come Szergedy, Zaremba, Sutskever, Bruna, Erhan, Goodfellow, Fergus (2013).

che l'immagine di partenza e quella modificata vengono identificate spesso come la stessa immagine – tali da far fallire il riconoscimento, come mostrato dall'immagine della figura 4.

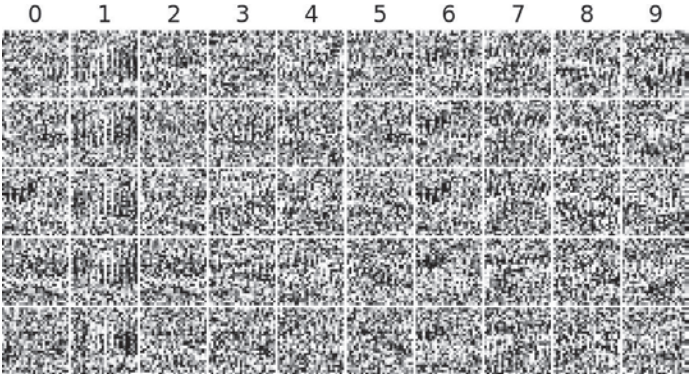

Figura 3. Immagini irregolari codificate che la rete neurale MNIST crede siano, con 99,99% di fiducia, le cifre 0-9. Ogni colonna è una classe di cifre, e ogni riga è il risultato dopo 200 generazioni di un percorso di evoluzione casuale e indipendente.
Fonte: NGUYEN A., YOSINSKI J., CLUNE J. (2015)

Perché dunque questi errori? Le immagini "sintetiche" create attraverso algoritmi evolutivi – comprese quelle contenenti rumore bianco – si basano su due procedimenti distinti: da un lato ci sono immagini codificate direttamente (*directly encoded*) attraverso una codifica indipendente e una ottimizzazione dei colori, secondo la codifica HSV[13], per ogni singolo pixel (figura 5).
Dall'altro le immagini vengono codificate per via indiretta, basata su pattern regolari, come già visto in precedenza

[13] La codifica dei colori HSV, similmente alle più note RGB (Red, Green, Blue) o CMYK (Ciano, Magenta, Yellow, Key Black), sta a indicare un metodo (additivo) di composizione per la rappresentazione in un sistema digitale. L'acronimo sta per Hue, Saturation, Value (tonalità, saturazione e valore).

(nella figura 2). Sia in un caso che nell'altro però queste immagini sono prive di significato.

(a) (b)

Figura 4. Esempi contraddittori per la rete QuocNet. Un classificatore per le automobili è stato addestrato senza fine-tuning. Gli esempi casualmente scelti a sinistra sono riconosciuti correttamente come automobili, mentre le immagini al centro non vengono riconosciute. La colonna a destra è il valore assoluto ingrandito della differenza tra le due immagini.

Fonte: SZERGEDY, ZAREMBA, SUTSKEVER, BRUNA, ERHAN, GOODFELLOW, FERGUS (2013)

Le Scienze Cognitive possono venire in aiuto alle reti neurali su questo aspetto specifico?

3. Regole e similarità[14] – in medicina e non solo

Gli esempi proposti in precedenza sembrano di fatto una delle possibili implementazioni di un problema cognitivo specifico: la capacità di categorizzare. Molte teorie si sono succedute in questo specifico settore delle Scienze Cognitive[15], ma un punto cruciale riguarda ciò che rende possibile il processo di (una corretta) categorizzazione.

[14] Similarità è una traduzione letterale dell'inglese *similarity*. Si è preferito mantenere l'assonanza con la lingua d'origine rispetto al più italiano somiglianza, perché in questa sede il termine ha una precisa valenza tecnica: la similarità è infatti da intendersi come elemento cognitivo necessario alla spiegazione di uno dei meccanismi responsabili della capacità di categorizzare.

[15] Per una rassegna mi permetto di rimandare a CELI (2008).

Una nozione fondamentale consiste nel riconoscere che le categorie sono composte da elementi similari, anche se molti ricercatori hanno scoperto che vi sono dissociazioni tra la categorizzazione e la similarità.

Barsalou (1985) ha dimostrato l'esistenza di categorie derivate da obiettivi (per esempio, la categoria dei possibili regali di compleanno) nelle quali gruppi altamente eterogenei sono messi insieme per uno scopo comune. Rips (1989) ha mostrato che certe proprietà di tipo naturale possono influenzare differentemente la classificazione e i giudizi di similarità.

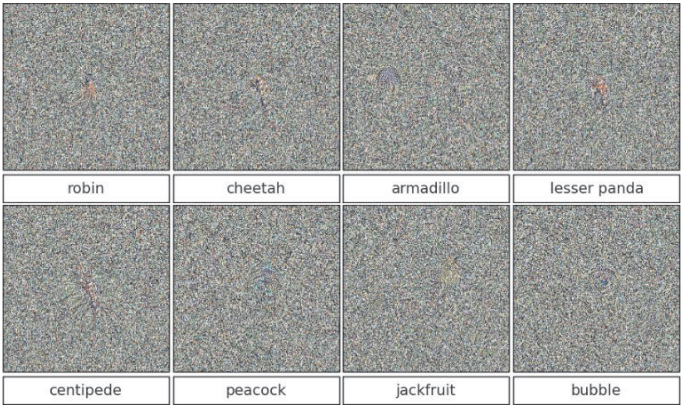

Figura 5. Codifica diretta delle immagini con ottimizzazione dei singoli pixel.
Fonte: NGUYEN A., YOSINSKI J., CLUNE J. (2015)

Sembrerebbe quindi intuitivamente che i giudizi di similarità siano fortemente dipendenti dalle proprietà percettive piuttosto che da altre proprietà, anche se le dissociazioni riscontrate sperimentalmente tra la categorizzazione e la similarità sono legate a singoli soggetti: in certi casi infatti alcune persone adottano un metodo analitico di risposta anziché fornire risposte su base intuitiva ed imme-

diata (Smith e Sloman, 1994). Questa dissociazione conduce un certo numero di ricercatori (per esempio Carey, 1985; Keil, 1989) a rigettare la similarità come base della categorizzazione, a favore della visione secondo la quale la categorizzazione dipende da teorie naif su certi domini. Questi ricercatori propendono per una spiegazione causale della categorizzazione, evidenziando gli aspetti soggiacenti che determinano l'appartenenza o meno a determinati insiemi e sappiamo che tale visione della categorizzazione generalmente contrasta con modelli a prototipi, ad esemplari e connessionisti, che assumono (tutti) la similarità come relazione critica per la formazione delle categorie.

In sostanza, Sloman e Rips (1998): (1) si mostrano d'accordo col fatto che la categorizzazione non può essere ridotta *in toto* ad una gradazione della similarità; (2) tale gradazione non può essere indipendente da contesto e (3) credono sia necessario eliminare la possibilità che tutto passi attraverso la similarità.

Alcuni dei meccanismi basati sulle regole sono necessari per spiegare (a) le nostre competenze nell'uso sistematico e produttivo del linguaggio e (b) il senso di certezza associata a qualche inferenza anche per domini non familiari. Ma vediamo più da vicino cosa vogliamo intendere con categorizzazione, e cerchiamo di stabilire in maniera più precisa quale sia il suo rapporto con regole e similarità.

3.1. *Il dermatologo di Smith*

È arrivato il momento di una più precisa definizione delle procedure di categorizzazione possibili. Una rassegna della letteratura sui concetti e sulla categorizzazione, suggerisce, come minimo, tre procedure distinte.

> Per decidere se un oggetto appartiene ad una certa categoria, si deve:
> - determinare se questo oggetto si attaglia a regole che definiscono la categoria (le regole che specificano le condizioni necessarie e sufficienti per l'appartenenza ad una categoria);
> - determinare la similarità dell'oggetto ad esemplari ricordati della categoria, oppure;
> - determinare la similarità dell'oggetto al prototipo della categoria[16].

Lavori seguenti[17] suggeriscono di aggiungere un'altra strategia alla lista: determinare se le caratteristiche dell'oggetto sono spiegate meglio attraverso una 'teoria' che soggiace alla categoria.

Prenderemo in considerazione solo le prime due procedure di classificazione a cui possiamo riferirci come ad "applicazione di regole" e "similarità con esemplari". Questa scelta è dettata dal fatto che esse sono tra le più discusse in letteratura e, come abbiamo potuto vedere, sono quelle maggiormente distinguibili sia in studi cognitivi che in studi neuropsicologici. A titolo di esempio[18] è possibile considerare la situazione in cui un dermatologo debba decidere se una lesione particolare della pelle sia imputabile alla (e quindi costituisca una prova della) malattia X.

Il dermatologo può intraprendere due strade; la prima – che riguarda l'applicazione di regole – ci dice che egli deve poter conoscere una regola del tipo "se la lesione ha un numero sufficiente di caratteristiche seguenti – forma ellittica, struttura irregolare, colore rosso scuro/marrone, ecc. – allora indica la malattia X". Se il dermatologo applica questa regola nel fare la sua diagnosi (categorizzazione), allora egli

---

[16] SMITH E MEDIN (1981).
[17] MURPHY E MEDIN (1985); SMITH, PATALANO, JONIDES (1998), p. 169.
[18] Esempio ispirato da BROOKS, NORMAN, ALLEN (1991).

> sarà impegnato nei processi che seguono:
> 1. avrà un'attenzione selettiva verso ogni attributo critico dell'"oggetto" (per esempio: la forma, la struttura e il colore della lesione);
> 2. per ognuno di questi attributi determinerà se l'informazione percettiva esemplifica il valore specificato nella regola (per esempio: "il colore è rosso scuro/marrone?"); e
> 3. 'mescolerà' le conseguenze del punto 2 per determinare la categorizzazione finale[19].

Questo modello schematico a tre stadi di applicazione delle regole è compatibile con numerose discussioni sul seguire una regola[20]. Il dermatologo può decidere di intraprendere una seconda strada – quella basata sulla similarità – in cui applica una comparazione tra la lesione che ha davanti e altre viste da lui precedentemente. In aggiunta alla conoscenza delle regole mostrate qui sopra, se il dermatologo ha visto anche molti pazienti probabilmente avrà memorizzato numerosi esemplari di malattie della pelle. Conseguentemente, egli può notare che la lesione che ha davanti agli occhi è molto simile agli esemplari immagazzinati della malattia X.

> Adesso la sequenza di processi include:
> - il recupero di esemplari immagazzinati (di varie categorie di malattie) che siano simili all' "oggetto"; e
> - la selezione di quella categoria i cui esemplari recuperati sono in qualche misura più simili all'"oggetto"[21].

Da notare l'importanza di un elemento che fino a questo momento abbiamo solo accennato: la memoria. Il fatto

---

[19] SMITH, PATALANO, JONIDES (1998), p. 170.

[20] Questa espressione evocherà elementi teorici di matrice wittgensteiniana, ma qui siamo in tutt'altro contesto.

[21] SMITH, PATALANO, JONIDES (1998), p. 170.

che il dermatologo possa effettuare una comparazione è legata a doppio filo al numero di volte che ha visto lesioni di quel genere e a come le ricorda. Ancora una volta, regole e similarità si compenetrino: le regole del primo caso contengono comunque al loro interno un processo di comparazione (in particolare la regola 2), mentre la similarità – elemento cardine del secondo modo di operare – è comunque strutturata e legata a un algoritmo, a una procedura e, in ultima istanza, a una regola.

3.2. *I problemi della similarità*

Da questa breve analisi emerge quel che Frank Keil rese evidente nella sua ricerca[22]: la similarità in sé non può essere intesa come unico costrutto utile e funzionale alla spiegazione dell'apprendimento nello specifico compito della categorizzazione. Al netto di una ulteriore distinzione tra oggetti naturali e oggetti fatti dall'uomo (*artifacts*) [23], Keil evidenziò il limite del solo uso della similarità con un esempio semplice ed efficace: se si dovesse utilizzare solo questo elemento per spiegare come le persone siano in grado di utilizzare il bianco, per esempio, questo avrebbe lo stesso peso nell'identificazione degli orsi polari e delle lavatrici[24]. Risulta evidente che così non può essere: mentre negli orsi polari il bianco è un attributo fondamentale alla sopravvivenza della specie – pur minacciata oggi da ben altri flagelli – e quindi ha valore causale, il bianco

[22] In bibliografia KEIL, SMITH, SIMONS, LEVIN (1998).

[23] Distinzione che in questa sede non affronteremo perché esula dall'argomentazione principale di questo progetto, ma sulla quale vi è un rinnovato interesse filosofico, come mostrato dal seminario di SOAVI (2017).

[24] Per un'analisi più specifica di questo argomento, mi permetto di rimandare a CELI (2004).

nelle lavatrici è in qualche modo del tutto casuale[25]. I due "bianchi" nominalmente appartenenti alla stessa categoria, hanno ruoli complementari (causale/casuale) nella spiegazione e questo inficia la possibilità di usare la sola similarità per spiegare la capacità di categorizzare.

Il dibattito scientifico sulla genuina autonomia della componente modellistico-teorica della categorizzazione all'interno delle Scienze Cognitive è stata messa in dubbio da più parti[26], ma la validità della distinzione qui proposta tra regole e similarità – e soprattutto una sua possibile implementazione in reti DL – resta, secondo noi, un valido suggerimento.

4. Conclusioni

La lacuna mostrata dall'uso della sola similarità come spiegazione per l'apprendimento di categorie nella cognizione umana, sembra avere qualche analogia con i fallimenti cui le reti neurali sono soggette per compiti di classificazione che oggi riteniamo semplici. Abbiamo visto che molte di queste soffrono di sovradeterminazione in certi casi e di sottodeterminazione in altri, impedendo – o limitando fortemente – l'uso delle stesse in contesti più ampi. Abbiamo accennato al fatto che l'apprendimento delle reti neurali viene realizzato sostanzialmente fornendo

[25] Per essere più precisi si tratterà senz'altro di una causalità di tipo differente che, pur non sapendo nulla sulla produzione industriale delle lavatrici, possiamo supporre sia da indicare in qualche risparmio di tipo economico nel processo di produzione di lamiere tutte ugualmente bianche. Una causalità che però definiremmo senz'altro più debole, nella considerazione che nulla vieta di avere lavatrici di un qualsivoglia altro colore. Queste "sopravvivono" nell'ambiente, a differenza di un orso che per sbaglio nasca con una pigmentazione della pelliccia diversa dal bianco.

[26] PRINZ (2002), ROGER, MCCLELLAND (2006).

moltissimi esempi di uno stesso oggetto in contesti molto diversi, in modo da affinare la capacità di riconoscimento.

Ciò che qui vogliamo ipotizzare è che questa strategia – che permette alle reti un apprendimento di fatto automatico – sembra trovare i suoi limiti proprio perché l'unico elemento di spiegazione è una "similarità senza ragione": il bianco per le lavatrici e gli orsi, ipotizzando una implementazione all'interno di una rete neurale, avrebbe dunque lo stesso peso, esattamente come hanno lo stesso peso le tecniche di codifica diretta (sui singoli pixel) e indiretta (sui pattern). Ma questo è insensato se tra due esempi dello stesso oggetto – quale che sia il livello di classificazione – non si dà alcuna regola soggiacente che possa identificare, e soprattutto distinguere in qualche modo, quali similarità (a uno qualsiasi dei livelli) abbiano senso e quali no. Uno strumento anche molto potente e duttile come la rete neurale è così destinato a fallire perché manca, a nostro avviso, di un ingrediente fondamentale di cui, nel corso di questo articolo, abbiamo discusso: l'uso di una o più regole. Usare solo queste ultime – come ci insegna la storia delle teorie concettuali – sarebbe ugualmente sbagliato, proprio perché comporterebbe una tassonomia probabilmente troppo rigida, con condizioni necessarie e sufficienti di inclusione in una determinata classe, da cui molti oggetti del mondo reale sfuggirebbero.

La sola similarità non basta e forse ciò che i cognitivisti hanno identificato a suo tempo come *Modello Ibrido* potrebbe essere visto – ammesso e non concesso che implementazione di quest'ultimo si possa attuare in una rete neurale – come un tentativo per superare la difficoltà evidenziata in queste pagine.

Riferimenti bibliografici

BARSALOU L. W. (1985), *Ideals, Central Tendency, and Frequency of Instantiation*, «Journal of Experimental Psychology: Learning, Memory and Cognition», 11, pp. 629-654.

BROOKS L. R., NORMAN G. R., ALLEN S. W. (1991), *Role of specific similarity in a medical diagnostic task*, «Journal of Experimental Psychology», 120 (3), pp. 278-287.

CAREY S. (1985), *Conceptual Change in Childhood*, Cambridge, MA: MIT Press, Massachusetts.

CELI L. (2004), *White bears and washing-machines: models (and theories) of concepts*, in Negrotti M. (ed.) «Yearbook of the Artificial II: Models in contemporary Science», Peter Lang Publisher, Berna, pp. 137-162.

CELI L. (2008), *La valigia di Aristotele. Concetti, prototipi, orsi bianchi e lavatrici*, Aracne Editrice, Roma.

CORDESCHI R. (2002), *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Kluwer Academic Publishers, Dordrecht.

FAWZI A., MOOSAVI-DEZFOOLI S.-M., FROSSARD P. (2016), *Robustness of classifiers: from adversarial to random noise*. In D. D. LEE, M. SUGIYAMA, U. V. LUXBURG, I. GUYON, & R. GARNETT (Eds.), *Advances in neural information processing systems 29* (pp. 1632–1640). Curran Associates, Inc.

FORNELL D. (2017), *How Artificial Intelligence Will Change Medical Imaging*, «itn. Imaging Technology News», all'indirizzo[27]:
https://www.itnonline.com/article/how-artificial-intelligence-will-change-medical-imaging.

[27] Tutti i riferimenti web sono stati controllati alla data del 6 agosto 2017.

KEIL F. (1989), *Conceps, Kinds and Conceptual Development*, Cambridge, MA: MIT Press, Massachusetts.

KEIL F. C., SMITH W. C., SIMONS J. D., LEVIN D. T. (1998), *Two Dogmas of Conceptual Empiricism: Implications for Hybrid Models of the Structure of Knowledge*, «Cognition», 65, pp. 103-135.

LECUN Y., BENGIO Y., HINTON G. (2015), *Deep learning*, «Nature», 521, pp. 436-444.

LOSANO M. G., (1991) *Storie di automi. Dalla Grecia classica alla Belle Epoque,* Einaudi, Torino.

LU J., SIBAI H., FABRY E., & FORSYTH D. (2017), *NO need to worry about adversarial examples in object detection in autonomous vehicles*, arXiv:1707.03501v1.

MURPHY G., MEDIN D. (1985), *The Role of Theories in Conceptual Coherence*, «Psychological Review», 92, pp. 289-316.

NGUYEN A., YOSINSKI J., CLUNE J. (2015), *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*. In *Computer Vision and Pattern Recognition (CVPR '15)*, IEEE.

PRINZ J. (2002), *Furnishing the mind-concepts and their perceptual basis*. Cambridge (MA): MIT Press.

RIPS L. J. (1989), *Similarity, Typicality, and Categorization*, in S. Vosniadou, Ortony A. (eds.), *Similarity and Analogical Reasoning*, NY: Cambridge University Press.

ROGERS T. T., MCCLELLAND J. L. (2006), *Semantic cognition - a parallel distributed processing approach*. Cambridge (MA): MIT Press.

SCHMIDHUBER J. (2015), *Deep learning in neural networks: An overview*, «Neural Networks», 61, pp. 85-117.

SENARATNE R., HARDY D., VANDERAA B., HALGAMUGE S. (2007), *Driver Fatigue Detection by Fusing Multiple*

Cues, in LIU D., FEI S., HOU Z., ZHANG H., SUN C. (eds.) *Advances in Neural Networks, 4th International Symposium on Neural Networks*, Nanjing, China, part II, pp. 801-809, Springer-Verlag, Berlin Heidelberg.

SLOMAN S., RIPS L. J. (1998), *Similarity as an Explanatory Construct*, «Cognition», 65, pp. 87-101.

SMITH E. E., PATALANO A. L., JONIDES J. (1998), *Alternative Strategies of Categorization*, «Cognition», 65, pp. 167-196.

SMITH E. E., MEDIN D. (1981), *Categories and Concepts*, Cambridge MA: Harvard University Press.

SMITH E. E., SLOMAN S. (1994), *Similarity- vs. Rule-based Categorization*, «Memory and Cognition», 22, pp. 377-386.

SOAVI M. (2017), *Inestricabili intrecci tra artefatti e natura*, seminario tenuto alla Scuola Estiva in Storia e Filosofia della Scienza «Scienza, tecnologia e società», Università di Padova, 17-19 luglio.

SZEGEDY C., ZAREMBA W., SUTSKEVER I., BRUNA J., ERHAN D., GOODFELLOW I., FERGUS R. (2013), *Intriguing properties of neural networks*, arXiv:1312.6199.

TAMBURRINI G. (2005), *Ethical Issues in Robotics, Bionics, and AI*. In WEBER J. (2008), *Techno-Ethical Case-Studies in Robotics, Bionics, and Related AI Agent Technologies* (Deliverable 5 of the EU-Project ETHICBOTS: *Emerging Technoethics of Human Interaction with Communication, Bionic and Robotic Systems*).

VANRULLEN R. (2017), *Perception science in the age of deep neural networks*, «Frontiers in Psychology», 8, 142, pp. 1-6.

ZHOU E., CAO Z., YIN Q. (2015), *Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not?*, arXiv:1501.04690v1.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130666
pp. 95-118 (marzo 2018)

Concettualizzare l'autoconsapevolezza corporea e le sue basi cognitive

Definizioni e tassonomie

di EDOARDO FUGALI[1]

1. Introduzione

Lo scopo che anima questo studio consiste in un tentativo di definizione concettuale dell'autoconsapevolezza – o certezza di sé – corporea, intesa come la forma minimale di autocoscienza implementata da processi di integrazione multisensoriale e motoria. Questo livello basilare prefigura le forme di autocoscienza di ordine superiore che richiedono risorse cognitive di natura concettuale e linguistiche più complesse ed elaborate. Tale obiettivo verrà perseguito nel quadro di un approccio integrato che tenga in debito conto sia le analisi vertenti sullo strato esperienziale del senso di sé corporeo proposte in sede di fenomenologia e filosofia della mente, sia le indagini sperimentali sviluppate nelle neuroscienze cognitive sui meccanismi che ne costituiscono la base funzionale, così come i relativi tentativi di concettualizzazione e classificazione, in buona parte tributari delle tassonomie già depositatesi nel senso comune. Si tenterà inoltre di ovviare alla mancanza di una definizione unitaria di autoconsapevolezza corporea, che sia condivisibile tra i vari ambiti disciplinari presi in esame.

[1] Università di Messina, Dipartimento di Scienze cognitive, psicologiche, pedagogiche e degli studi culturali. E–mail: efugali@unime.it

Secondo la mia tesi di fondo, coscienza e autocoscienza non sono emanazioni di una mente disincarnata, ma sono predelineate nella loro genesi già dall'autoconsapevolezza corporea, qui definita come la certezza irrefutabile di essere il latore delle proprie sensazioni corporee (*senso di proprietà*) e l'iniziatore dei propri movimenti volontari (*senso di agentività*). Come gli stati di coscienza di ordine superiore, anche quelli vertenti sul proprio sé corporeo godono della proprietà del riferimento a sé esibita da ogni vissuto d'esperienza, ma se ne differenziano nella misura in cui fanno capo a un tipo di consapevolezza preriflessiva, non tematica e non linguisticamente articolata. Entrambe le specificazioni dell'autocoscienza corporea presuppongono, sia pure in misura variabile, la dimensione fenomenologica del corpo vissuto, ossia il centro prospettico e l'istanza di localizzazione da cui si irradiano i nostri stati mentali e le nostre azioni. Il contesto teorico retrostante a questo tentativo è definito dalla scienza cognitiva *embodied* nelle sue intersezioni con la fenomenologia, di cui essa riprende la caratterizzazione del corpo vivo come strato costitutivo del soggetto e della cognizione.

Nel § 2 analizzerò i componenti di base dell'autoconsapevolezza corporea, cioè il senso di proprietà e il senso di agentività al fine di evidenziarne le proprietà strutturali, ossia l'essere esperiti alla prima persona, la localizzazione di sé, l'immunità da errori di autoidentificazione e l'assenza di prospetticità. Nel § 3 mi soffermerò sulle strutture rappresentazionali che li sorreggono e rinviano a due differenti ordini categoriali, cioè lo schema corporeo e l'immagine corporea, per evidenziarne la sostanziale corrispondenza con le due dimensioni del corpo vissuto e del corpo oggetto individuate in fenomenologia e sottoporre a verifica critica l'adeguatezza dei concetti e

delle categorizzazioni dei modi di esperienza del corpo invalsi nelle scienze cognitive contemporanee.

2. *Et in carne mea*. L'autoconsapevolezza corporea e le sue componenti

A dispetto della sua apparente unitarietà fenomenologica, l'autoconsapevolezza corporea si presta a un novero estremamente variegato di definizioni, categorizzazioni e tassonomie che ne evidenziano unilateralmente, spesso a seconda delle preferenze teoriche degli autori, l'uno o l'altro degli aspetti che contribuiscono a costituirla (Fugali, 2012; 2016). In prima approssimazione sono da individuare sotto questo riguardo almeno due approcci, che in scienza cognitiva e in filosofia della mente rinviano rispettivamente al paradigma tradizionale basato sulla preminenza di rango della rappresentazione e a quello *embodied* incentrato invece sul primato dell'azione, di cui il corpo rappresenta il veicolo fondamentale.

Tra i sostenitori del primo approccio si distinguono coloro che, come Bérmudez (2011), propongono una concezione di stampo esplicitamente deflazionario. Bérmudez sostiene che il corpo non è tanto l'oggetto di una rappresentazione dedicata, caratterizzata da uno specifico marchio qualitativo e fenomenologico, quanto il termine di riferimento di giudizi alla prima persona di natura concettuale. Per quanto formulati in base a fatti relativi alle sensazioni corporee, che sotto il profilo qualitativo non si distinguono da quelle vertenti su oggetti esterni, tali giudizi sono logicamente indipendenti da essi.

I rappresentazionalisti propendono invece per una visione più liberale che contempla anche l'esistenza di genuine rappresentazioni corporee di origine percettiva, e si

avvalgono di differenti criteri di classificazione, succedutisi a partire dalla distinzione classica introdotta in letteratura da Head e Holmes (1911/12) tra schema corporeo e immagine corporea di cui si discuterà più diffusamente nel § 3. A questa si affianca un'ulteriore distinzione tra rappresentazioni a breve termine, ossia percezioni corporee consce, e rappresentazione a lungo termine – attribuzioni valutative, credenze ecc. che fungono a livello disposizionale (O'Shaughnessy, 2000 e 2008) –, sostanzialmente coincidente con quella tra rappresentazioni on–line e off–line proposta da G. Carruthers (2008).

Infine, i fautori del paradigma *embodied* riconducono il carattere di fondamentale unitarietà dell'autoconsapevolezza corporea e la sua apparente compattezza fenomenica alla categoria del corpo in azione, rinvenendone i fattori causali nel gioco di contingenze sensorimotorie su cui si fonda l'inestricabile legame tra percezione e azione. In conformità alla posizione di preminenza di cui questa gode rispetto alla prima, si individua nel corpo, inteso come sistema dell'"io–posso" (Husserl, 1952) e punto focale di irradiazione di un'intenzionalità motoria che precede quella coscienziale (Merleau–Ponty, 1945), l'ingrediente fondamentale del senso di autoconsapevolezza corporea, che non richiede l'interposizione di meccanismi rappresentazionali funzionali a ritrarne le caratteristiche metriche e in genere quelle che lo rendono assimilabile a un qualunque oggetto esterno. Centrale sotto tal riguardo è la distinzione, introdotta da Husserl, tra la dimensione del corpo vissuto in prima persona (il *Leib*) e quella del corpo materiale (*Körper*), che almeno in parte coincide, come si evidenzierà più dettagliatamente in seguito, con quella tra schema e immagine corporea (§ 3).

Questa disparità di opinioni è il sintomo della difficoltà di rinvenire una definizione univoca e ampia almeno quan-

to basta per tenere assieme sia le presupposizioni sedimentatesi nell'esperienza comune, che sono oggetto diretto delle considerazioni fenomenologiche, sia le acquisizioni maturate in sede di analisi concettuale e di (neuro)scienza cognitiva. Diversi autori, tra cui sono da annoverare soprattutto neuroscienziati, sono stati indotti a rinunciare a tale impresa e a denegare al senso di certezza corporea ogni effettiva consistenza, fino ad affermarne senza mezzi termini l'illusorietà: l'esperienza del nostro corpo, a dispetto dei suoi caratteri di familiarità e ubiquità, sarebbe in realtà un costrutto evanescente e precario, che elegge a proprio modello profili e fattezze del corpo fisico ed è costituito da una molteplicità di componenti cognitive e neuronali, dissociabili grazie a procedure psicometriche e psicofisiche (Stamenov, 2005; Longo, Haggard, 2012).

Quest'assunto equivale all'avallo di una posizione di sapore schiettamente eliminativista, nel momento stesso in cui si riconosce esclusiva legittimità al piano concettuale dell'indagine neuroscientifica, che dovrebbe escludere dal proprio ordine di considerazione ogni definizione che rinvii alle categorie della psicologia del senso comune. I suoi sostenitori sono certo disposti a concedere qualche margine di plausibilità all'indagine fenomenologica e all'analisi concettuale delle forme in cui si articola l'esperienza corporea, e attribuiscono a esse l'indubbio merito di aver fornito descrizioni altamente ricche e sofisticate delle molteplici forme in cui l'autoconsapevolezza corporea si articola sul piano esperienziale. Tuttavia, si tratta di metodi pressoché privi di rilevanza operazionale ai fini di una ricerca empirica condotta in base a criteri metodologici rigorosi: si spiega così l'esigenza di sottoporre la coscienza del Sé corporeo a una procedura analitica che ne evidenzi le componenti dissociabili e sia in grado di porre i ricercatori nelle condizioni di avanzare delle previsioni passibili

di misurazione diretta e gestibili in sede di indagine sperimentale (Longo *et al.*, 2008).

Anche alla luce di queste riserve, che condivido peraltro solo in parte, non credo sia del tutto preclusa la possibilità di pervenire a una definizione unitaria dell'autoconsapevolezza corporea atta a catturarne la varietà di aspetti nella loro reciproca connessione. Provo a proporne una formulazione: l'autoconsapevolezza corporea nel suo nucleo minimale è la certezza immediata e irrefutabile da parte di chi la intrattiene d'essere il latore delle proprie sensazioni corporee e l'iniziatore dei propri movimenti volontari, che si irradiano prospetticamente a partire dal "qui" e "ora" in cui il proprio corpo è localizzato.

Alcune definizioni analoghe a questa, tra quelle suggerite nella letteratura più recente, sono state proposte da Borghi e Cimatti (2010) e da de Vignemont (2011). Per i primi, l'autoconsapevolezza corporea coincide con la sensazione basilare di essere un corpo che agisce e percepisce, mentre de Vignemont la equipara alla certezza immediata e interna dei propri movimenti, veicolata da rappresentazioni di natura propriocettiva e motoria. Dal canto loro, Gallese e Sinigaglia (2010 e 2011) ritengono di poterla circoscrivere alla certezza di sé in quanto soggetto corporeo di azioni, partendo da un'accezione piuttosto radicale dell'approccio *embodied* che insiste sul primato dell'azione e sul ruolo pressoché esclusivo svolto dai processi motori. Secondo Metzinger (2010) l'autoconsapevolezza corporea rimanda alla struttura minimale del *core self*, ossia un modello rappresentazionale del proprio corpo, elaborato dal cervello sulla base di processi di integrazione multisensoriale e motoria. Queste definizioni, tuttavia, colgono solo in parte la ricchezza degli aspetti e delle componenti di cui l'autoconsapevolezza corporea consiste, giacché ne evidenziano soltanto alcune

– l'azione, il senso di proprietà, il senso di agentività o la componente motoria.

L'articolazione interna del fenomeno dell'autoconsapevolezza corporea potrà forse emergere più chiaramente se proviamo ad analizzarne le componenti, le proprietà caratteristiche e le risorse cognitive a cui attinge. Si tratterà poi di capire quali tra questi aspetti spettino solo a essa e quali siano condivisi con le forme più evolute e cognitivamente complesse di autocoscienza. A tal fine ritengo necessario tornare sulla definizione che ho appena provato a tratteggiare.

Il dato immediato in cui veniamo a imbatterci nel momento in cui guardiamo da vicino a questo fenomeno è che il nostro corpo ci si manifesta al tempo stesso come capacità di sensazione e potere d'azione. In altri termini, riguardo ad esso intrattengo la consapevolezza indubitabile e immediata di essere il portatore di ogni mia sensazione e l'istanza prima da cui si originano i miei movimenti volontari. Questi due aspetti rinviano a due componenti del sé personale individuate nella letteratura più recente in sede di filosofia della mente e scienze cognitive, ossia il *senso di proprietà* (corporea) e il *senso di agentività*. Se il primo è definito come la sensazione o il sentimento di appartenenza del proprio corpo che qualifica l'esperienza che ne faccio come dall'interno e che contrassegna questo corpo fisico che mi capita di essere in quanto "il mio", il secondo investe invece la certezza altrettanto indefettibile di essere l'autore delle proprie azioni consapevoli e volontarie (Gallagher, 2000, 2005). Una tassonomia più recente distingue nelle due componenti un livello base attinente alla sensazione e due superiori attinenti rispettivamente alla concettualizzazione semantico–linguistica operante nei giudizi di proprietà e di agenzia e all'attribuzione metarappresentazionale e riflessiva di questi aspetti

dell'autoconsapevolezza corporea a sé e ad altri, che suppongono la dimensione sociale e normativa (Synofzik, Vosgerau, Newen, 2008).

Dal punto di vista funzionale, le due componenti differiscono poi per il fatto che il senso di agentività induce una forma maggiormente globale e coerente di certezza propriocettiva rispetto al senso di proprietà, nonché per la loro relazione di fondazione: attribuire a me stesso il ruolo di iniziatore di un'azione implica necessariamente essere consapevole delle membra corporee che impiego nell'eseguire i movimenti che la realizzano. Nulla infatti mi impedisce di continuare a mantenere il mio senso corporeo di proprietà anche in assenza di movimenti volontari (Tsakiris, Schütz–Bosbach, Gallagher, 2007). Nell'esperienza quotidiana, senso di proprietà e senso di agentività concorrono a dar vita alla certezza di sé corporea e a impregnare di sé tutte le nostre azioni e movimenti, tanto che le rispettive fenomenologie sono quasi indistinguibili, data l'immediatezza con cui viviamo l'"esser sempre qui" del corpo. Tra le due componenti è tuttavia da rilevare una fondamentale dissimmetria, come risulta da una serie di esperimenti neuropsicologici. Il *setting* di laboratorio in essi proposto ha reso possibile studiare mediante opportune manipolazioni il senso di agentività, che accompagna il solo movimento attivo implicato nelle azioni volontarie separatamente da quello di proprietà, esperito invece sia durante i movimenti attivi sia durante quelli passivi e involontari (Tsakiris, Haggard, 2005; Tsakiris, Prabhu, Haggard, 2006; Tsakiris, Longo, Haggard, 2010). È inoltre emerso come il senso di agentività sia fasico e intermittente rispetto quello di proprietà, che viceversa accompagna di continuo la consapevolezza del soggetto che lo intrattiene (Tsakiris, Fotopoulou 2013).

In seconda battuta, l'evidenza a cui non possiamo sot-

trarci nel considerare la relazione coscienziale che intratteniamo con le nostre sensazioni corporee e i nostri movimenti volontari è che essi recano indelebilmente impresso il contrassegno dell'appartenenza al soggetto che prova qualcosa a riguardo. Sono già da sempre consapevole delle *mie* sensazioni corporee e delle *mie* azioni in quanto mie, ossia in quanto sperimentate a partire dalla specifica prospettiva alla prima persona che mi consente di riferire a me stesso ogni mia esperienza, di qualunque natura essa sia. Terzo punto, questa prospettiva è invariabilmente localizzata nella porzione dello spazio–tempo che il mio corpo di volta in volta occupa: tutto quanto vi rientra si colloca all'interno di un sistema di coordinate egocentriche, coestensivo al corpo e allo spazio peripersonale che lo circonda, che non è esso stesso di natura prospettica.

Ritornerò poco oltre ad analizzare nel dettaglio i meccanismi cognitivi che sorreggono il funzionamento di questi componenti. Per ora intendo soffermarmi sugli aspetti caratteristici che l'autoconsapevolezza corporea condivide con altre forme di autocoscienza e su quelli che la contraddistinguono in quanto tale, nonché sulle molteplici fonti di informazione che contribuiscono alla sua origine.

Anzitutto l'autoconsapevolezza corporea costituisce il nucleo essenziale e minimale del nostro Sé prima di ogni sua ulteriore specificazione (Gallese, Sinigaglia, 2010) e si distingue dalle forme più elaborate di coscienza di sé – quali l'autocoscienza riflessiva e quella narrativa – per la sua natura non concettuale e pre–riflessiva. Sotto il profilo fenomenologico questa proprietà si riverbera sul suo specifico modo di manifestazione, ossia sul suo carattere recessivo: di norma infatti il corpo non è un oggetto tematico di attenzione, dato che nel corso consueto delle nostre azioni siamo sempre concentrati su oggetti e situazioni che trascendono la nostra corporeità.

Ciò non significa ovviamente che la certezza corporea si dilegui fino a essere riassorbita in un processo inconscio e subpersonale, ma solo che essa regredisce ai margini della nostra attenzione consapevole. È pur vero che il corpo condivide con altri oggetti questa caratteristica, dato che possiamo fare convergere a piacimento la nostra attenzione su questi o quegli oggetti o le loro caratteristiche salienti a discapito di altri. D'altra parte, solo il corpo tra tutti gli altri oggetti si manifesta per lo più in modo recessivo e trasparente, giacché costituisce il centro di irradiazione delle nostre capacità sensoriali e dei nostri poteri di azione, anche se sotto determinati aspetti può essere esperito come qualunque altro oggetto intenzionale.

Questo ci rimanda a un'ulteriore proprietà che caratterizza in via esclusiva il corpo tra tutti gli oggetti del mondo, dato che esso da una parte è condizione soggettiva di possibilità di ogni sensazione e azione, dall'altra è assimilabile a qualsiasi altro oggetto fisico. La duplicità funzionale del corpo si ripercuote al contempo sul modo in cui ne facciamo esperienza, dato che possiamo sia sentirlo "da dentro" grazie a un'apprensione immediata e diretta veicolata dalle sensazioni somatiche (enterocettive), sia al contempo guardare ad esso dall'esterno come a un oggetto fisico e biologico per mezzo di percezioni esterocettive e di atti cognitivi di ordine superiore, come riflessioni, credenze, giudizi o valutazioni di tipo estetico. In breve, solo il nostro corpo si fenomenizza anche dall'interno e può dunque costituirsi come corpo soggettivo.

Il singolare fenomeno del *touchant/touché*, su cui si sono diffuse le analisi descrittive di Husserl (1952) e Merleau–Ponty (1945), esprime con esemplare evidenza il doppio canale d'accesso al nostro corpo (cfr. Tsakiris, Haggard, 2005; de Vignemont, 2010). Tale peculiarità investe poi il carattere di privatezza degli aspetti fenomeno-

logici inerenti all'esperienza del Sé corporeo. Le determinazioni qualitative dei sensi esterni che mettono capo a percezioni d'oggetti – e che si indirizzano al corpo in quanto oggetto – sono almeno in una certa misura pubblicamente condivisibili, come accade per quelle visive, auditive e tattili. Diverso è invece il caso delle sensazioni somatiche, che sembrano godere di uno statuto di privatezza alla seconda potenza, essendo circoscritte a una dimensione del tutto inaccessibile a chiunque non ne sia il portatore (Bermùdez, Marcel, Eilan, 1995; Evans, 1982).

Sembra dunque, alla luce di quanto detto, che la certezza di sé corporea investa nella sua peculiarità più l'automanifestazione soggettiva del corpo che le proprietà da esso condivise con tutti gli altri oggetti. Il maggior grado di privatezza e intimità esibito dalle sensazioni somatiche e, in generale, da tutti i modi soggettivi di manifestazione del Sé corporeo è inestricabilmente connesso a due ulteriori proprietà dell'autoconsapevolezza corporea e potrebbe forse costituirne la base esplicativa, ossia l'immunità da errori di autoidentificazione e l'assenza di prospetticità.

Quanto al primo aspetto, un tratto distintivo che accomuna tutte le manifestazioni soggettive del nostro corpo è l'apprensione diretta e immediata di tali stati in quanto riferiti a noi stessi senza che sia necessario operare ulteriori accertamenti[2] (Bermùdez, 1998, 2011; Brewer, 1995; Cassam, 1995, 2011; Dokic, 2003; Evans, 1982). Posso certo ingannarmi sull'appartenenza a me stesso di una parte corporea oggetto di una percezione esterna: esemplare a tal proposito è l'illusione della mano di gomma, prodotta

[2] Vanno escluse alcune situazioni limite come il senso di sorpresa che talora proviamo quando ci capita di vedere all'improvviso la nostra immagine riflessa.

in un celebre quanto elegante esperimento (Botvinick, Cohen, 1998) che ha dato l'avvio a innumerevoli variazioni. Ai soggetti sperimentali veniva nascosta la mano vera mentre veniva loro mostrato un arto di gomma che ne riproduceva le fattezze. Nel momento in cui la mano vera e la finta erano sottoposte a stimoli tattili sincroni, essi localizzavano nella seconda le sensazioni tattili e propriocettive provate. Viceversa, sono consapevole in modo infallibile che sono proprio io a provare questa o quella sensazione corporea, ferma restando l'eventualità che possa ingannarmi sul suo contenuto, sulla sua natura o sulla sua localizzazione.

Riguardo all'assenza di prospetticità, il corpo sentito dall'interno non si manifesta lungo un decorso di aspetti parziali come avviene per gli oggetti esterni, che sono percepiti a distanza variabile e secondo la loro collocazione da un determinato punto di osservazione in modo sempre incompiuto, dato che mostrano solo un profilo per volta, ma si offre simultaneamente nella sua interezza. Resta tuttavia possibile che ciò avvenga secondo differenti gradi di intensità, in conformità alla fenomenologia recessiva dell'autoconsapevolezza corporea. In altre parole, per la certezza corporea soggettiva vige non la coppia oppositiva manifesto/nascosto, ma quella figura/sfondo, dato che nella percezione somatosensoriale certe parti corporee o certe specifiche sensazioni possono imporsi con un maggiore grado di evidenza rispetto ad altre che rimangono tuttavia oggetto di una consapevolezza marginale. Inutile precisarlo ancora, ciò non vale per il corpo oggetto della percezione esterna, sebbene i vincoli anatomici imposti dagli organi di senso modulino anche la nostra prospettiva visiva sul corpo. Questa si contraddistingue per la sua relativa invariabilità e la sua incompiutezza: non posso infatti percepire il volto e le parti posteriori del mio corpo se non ricorren-

do a specchi, telecamere o altri artefatti.

3. Il corpo agito e il corpo rappresentato

La duplicità dei modi d'accesso al corpo e dell'esperienza che ne abbiamo si spiega dunque in base ai differenti canali cognitivi che ci consentono di rappresentarlo. Il corpo è un costrutto multimodale in cui convergono materiali informazionali di funzione e provenienza eterogenee. Questo può forse contribuire a spiegare la sovrabbondanza di tassonomie e di modelli classificatori in cui si imbatte chiunque getti anche solo uno sguardo distratto sulla letteratura sull'argomento (v. *supra*, § 2). Possiamo suddividere sommariamente queste informazioni in un livello base e un livello di ordine superiore: nel primo rientrano sensazioni somatiche e rappresentazioni somatosensoriali relative alla superficie cutanea e alle parti del corpo, mentre il secondo consiste di percezioni di alto livello del corpo e degli oggetti con cui esso viene in contatto (somatopercezione) e di conoscenze astratte, credenze, attitudini, rappresentazioni affettive e attribuzioni di valore estetico e sociale (Longo, Azañón, Haggard, 2010).

I meccanismi rappresentazionali sottesi all'esperienza della nostra corporeità possono essere ricondotti a due strutture integrate note sotto le diciture di *schema corporeo* e *immagine corporea*. Questa terminologia si diffonde a partire dall'introduzione della tassonomia di Head e Holmes (1912), che individua tre tipi di rappresentazioni corporee. Lo *schema posturale* rileva in tempo reale la posizione delle membra e funge da istanza di controllo per l'esecuzione dei movimenti corporei. Lo *schema superficiale* si incarica di localizzare gli stimoli sensoriali sulla superficie della pelle. L'*immagine corporea* include inve-

ce le rappresentazioni consapevoli del corpo e delle sue parti. Nello schema corporeo rientrano dunque come suoi sub-componenti i primi due membri della tripartizione. Secondo una definizione più recente (Gallagher, 2005), ispirata alla sua caratterizzazione fenomenologica come struttura globale sottesa all'apprensione del corpo in quanto proprio e all'esercizio dell'intenzionalità motoria (Merleau–Ponty, 1945), lo schema corporeo consiste di un set di capacità sensori–motorie impiegate nei processi subpersonali, modulari e automatici preposti all'esecuzione e al controllo dell'azione. Sotto questo titolo sono da includere informazioni di natura tattile, propriocettiva, cinestesica e vestibolare.

Nell'immagine corporea confluiscono viceversa tutte le rappresentazioni personali, intenzionali e consapevoli del proprio corpo non strettamente funzionali al compimento dell'azione. Queste si generano a livello riflessivo grazie all'apporto congiunto di informazioni provenienti da tutti i canali sensoriali, laddove a rivestire un ruolo preponderante è tuttavia la modalità visuale. Tali costrutti comprendono non solo percezioni on–line, ma anche rappresentazioni a lungo termine che si sedimentano in attitudini disposizionali, come conoscenze concettuali e semantiche, credenze, disposizioni affettive e valutative. Pur nella loro disparata provenienza e modalità, queste rappresentazioni sono accomunate dal fatto di vertere intenzionalmente sul corpo in quanto oggetto, che viene appreso come ogni altro nell'avvicendarsi di una successione di aspetti parziali senza dar luogo a una rappresentazione globale e olistica, a differenza di quanto avviene nello schema corporeo.

In che modo si combinano ora le rappresentazioni dello schema e dell'immagine corporea nel dar vita al senso di proprietà e di agenzia? Anzitutto è da rilevare come una distinzione netta tra schema corporeo e immagine corpo-

rea richieda una riflessione supplementare di ordine concettuale, dato che, come nel caso del senso di proprietà e di agenzia, essi concorrono in modo strettamente congiunto a strutturare la nostra consapevolezza e il nostro agire. Ciò fa sì che nell'esperienza normale i loro confini siano molto più sfumati di quanto non possa apparire a prima vista, soprattutto per quanto riguarda il senso di proprietà.

È un'acquisizione ormai consolidata e corroborata da numerose evidenze sperimentali, come lo studio delle patologie dello schema e dell'immagine corporea (de Vignemont, 2010) e gli esperimenti sull'illusione della mano di gomma cui si è accennato sopra, che il senso di proprietà consti di due specifiche tipologie di informazioni. La prima comprende sensazioni afferenti di natura visiva, tattile, cinestesica e propriocettiva che si offrono in tempo reale e fungono a livello bottom–up. Nella seconda sono da annoverare rappresentazioni cognitive off–line preesistenti e permanenti (di natura visiva, propriocettiva, affettiva ecc.) che modulano le informazioni afferenti in direzione top-down. Nessuna delle due componenti è sufficiente da sola a produrre il senso di proprietà, mentre lo è l'apporto delle differenti modalità sensoriali anche in assenza del senso di agentività (de Vignemont, 2007; Tsakiris, Schütz-Bosbach, Gallagher, 2007).

Nella genesi della fenomenologia del senso di proprietà corporea rifluiscono dunque informazioni sensorie provenienti tanto dallo schema corporeo, quanto dall'immagine corporea. Direttamente implicati nella genesi del senso di agentività sono invece soltanto i comandi motori efferenti che precedono l'azione e traducono in movimento effettivo l'intenzione motoria e gli input sensoriali della copia efferente di feedback, ricalcanti le sensazioni afferenti che a livello bottom–up fungono da materiale grezzo per il senso di proprietà (Frith, Blakemore, Wolpert, 2000). Più

che con strutture rappresentazionali, qui abbiamo a che fare propriamente con eventi cinestesici, tattili e propriocettivi. La natura caratteristica dei processi cognitivi sottesi al senso di agentività trova un corrispettivo nella sua fenomenologia "sottile", per cui il corpo non è tanto l'oggetto di una certezza tematicamente indirizzata verso un correlato oggettuale, quanto una struttura trasparente e preriflessiva che retrocede sullo sfondo del corso globale d'azione a cui principalmente mira la nostra attenzione consapevole.

Alla differenza funzionale tra i meccanismi cognitivi preposti alla genesi della coscienza del sé corporeo fa dunque riscontro la fondamentale duplicità del modo in cui il nostro corpo si manifesta nella nostra esperienza ordinaria. Da una parte troviamo il corpo come soggetto, ossia il corpo che noi stessi "siamo" transitivamente e viviamo "dall'interno" che, pur fungendo di continuo alle spalle della nostra esperienza del mondo come sfondo unitario, si sottrae per lo più alla nostra consapevolezza diretta; dall'altra il corpo assimilato ad ogni altro oggetto intenzionale, che si manifesta alla nostra percezione solo attraverso scorci parziali.

Il corpo soggettivo, designato in fenomenologia col termine *corpo vivo* o *corpo proprio* (*Leib*) (Husserl, 1952; Merleau-Ponty, 1945), è un sistema unitario e integrato di organi di percezione e di movimento che costituisce il centro di irradiazione del senso di proprietà e di agenzia e si sorregge essenzialmente sulle risorse informazionali offerte dallo schema corporeo, in particolare i comandi motori efferenti e le sensazioni tattili interne e propriocettive. In modo analogo allo schema corporeo, il corpo vivo è una struttura olistica e globale aggiornata di continuo che accompagna in modalità on–line ogni vissuto e ogni azione senza mai venir meno del tutto. Alla costituzione del corpo

oggetto (*Körper*) concorrono invece tutte le informazioni percettive multimodali (ferma restando la predominanza di quelle visive), nonché le rappresentazioni concettuali, affettive e valutative preesistenti in modalità off–line comprese sotto il titolo dell'immagine corporea. Il corpo oggetto coincide dunque in via quasi esclusiva col versante del senso di proprietà corporea limitatamente alle modalità sensoriali esterocettive, dato che le capacità motorie implicate dal senso di agentività svolgono un ruolo marginale, consistente nel conferire unitarietà alle rappresentazioni parziali veicolate dall'immagine corporea.

4. Considerazioni conclusive

Proprio il carattere elusivo e ubiquitario dell'autoconsapevolezza corporea, che rileva dalla circostanza che esso accompagna come una sorta di sfondo silente ogni nostra esperienza, rende problematica un'indagine sui meccanismi cognitivi che la implementano, così come una ricognizione delle strategie di concettualizzazione e di categorizzazione a cui si fa ricorso tanto sul piano basilare del senso comune, quanto su quello più elaborato della descrizione fenomenologica e delle strategie esplicative intraprese dalle scienze cognitive.

Al primo livello sembrano passibili di concettualizzazione soltanto gli aspetti dell'autoconsapevolezza corporea che si dirigono intenzionalmente verso il corpo oggetto e che rientrano nell'immagine corporea: in questo senso, il corpo è individuato non solo da proprietà metriche, spaziali e semantiche relative alla sua configurazione e alle sue parti, che esso condivide con altri oggetti materiali, ma anche da attribuzioni valoriali di ordine estetico, pratico ed etico. Qui il senso di appartenenza si traspone solo per via

derivativa a partire dall'esperienza primaria del corpo vissuto nell'azione da cui esso trae origine. Come tale, nel corso delle nostre interazioni pratiche con l'ambiente, il corpo è una sorta di dato sempre superato in vista del contesto progettuale che esse pongono in essere (Sartre, 1943), sebbene possa giungere a manifestazione in certe esperienze cruciali, quali quella del *touchant/touché* menzionata in precedenza.

Le descrizioni fenomenologiche si trovano dunque di fronte al problema di selezionare in modo rigoroso, tra le molteplici esperienze percettivo-sensoriali che accomunano l'apprensione del corpo proprio a quella del corpo oggetto, quelle che elettivamente si indirizzano al primo, ossia la consapevolezza dell'iniziazione dei movimenti volontari (l'"io posso" di Husserl e l'intenzionalità motoria di Merleau–Ponty), le sensazioni del tatto interno e quelle propriocettive. Tanto le une quanto le altre rientrano in una struttura integrata e olistica che a dispetto della sua recessività si presta ad essere analizzata tramite gli strumenti dell'analisi fenomenologica.

Dal canto loro, le scienze cognitive hanno individuato in modo più dettagliato i processi rappresentazionali sottesi alle differenti modalità di esperienza del corpo, non senza tuttavia sottrarsi alle insidie di un approccio scompositivo e parcellizzante, spesso affetto da una sovrabbondanza di modelli concettuali che lascia in secondo piano i caratteri strutturali dell'esperienza del corpo – punto di vista alla prima persona, autolocalizzazione, immunità da errori di autoidentificazione, direzione dall'interno e assenza di prospetticità. Rispetto agli approcci più tradizionali di matrice rappresentazionalista, quello *embodied* ha se non altro il merito di raccordarsi alla tradizione fenomenologica nel prendere sul serio l'esperienza soggettiva quale base per l'elaborazione di modelli concettuali più aderenti a es-

sa e nell'individuare nell'azione la categoria descrittiva ed esplicativa fondamentale.

Alla luce delle analisi svolte sinora, credo di poter affermare che l'autoconsapevolezza corporea e le sue componenti condividano con le forme di autocoscienza di livello superiore molti più aspetti di quanto comunemente non si sia indotti a credere. Inoltre, come queste ultime, essa può essere sottoposta a strategie d'indagine in grado di evidenziare la molteplicità delle componenti soggiacenti alla sua genesi e le modalità in cui essa viene concettualizzata tanto sul piano fenomenologico dell'esperienza comune, quanto su quello delle categorizzazioni invalse negli ambiti disciplinari che la eleggono a oggetto.

La modalità d'accesso epistemico ed esperienziale ai propri stati interni – vale a dire, la prospettiva alla prima persona e la proprietà a questa connessa dell'immunità da errori di autoidentificazione – non spetta infatti soltanto agli enunciati-io (Wittgenstein, 1958), in cui il pronome "io" occorre in funzione soggettiva e alle forme più sofisticate di consapevolezza di sé. Anche le sensazioni somatiche, che si contraddistinguono rispetto a quelle esterocettive per la loro provenienza dall'interno e danno vita allo strato dell'esperienza del corpo come soggetto, condividono le medesime proprietà. Il tratto differenziale tra i due livelli di autoconsapevolezza consiste dunque nel fatto che quella corporea è atematica, preriflessiva e non linguisticamente articolata, almeno per quanto attiene al versante del corpo proprio.

A ben vedere, tuttavia, se è vero che il linguaggio è, prima che l'involucro di un senso evanescente, una tecnologia o un'estensione corporea, come sostengono gli esponenti delle varianti più radicali dell'approccio *embodied* (Lakoff, Johnson, 1999), anche le forme più astratte e sofisticate di autoconsapevolezza sono riconducibili alla di-

mensione del corpo. Affrontare tale questione nella dovuta ampiezza richiederà tuttavia uno studio espressamente dedicato che esula dai limiti di questo contributo.

Riferimenti bibliografici

Bermùdez, J.L., (2011) *Bodily awareness and self-consciousness*, in Gallagher, 2011, pp. 157–179.

Bermùdez, J.L, Marcel, A., Eilan, N. (a cura di), (1995) *The body and the self*, Cambridge University Press, Cambridge (MA).

Borghi, A.M., Cimatti, F., (2010) *Embodied cognition and beyond: Acting and sensing the body*, «Neuropsychologia», 48:763–773.

Botvinick, M., Cohen, J., (1998) *Rubber hand 'feels' touch that eyes see*, «Nature», 391:756.

Brewer, B., (1995) *Bodily awareness and the self*, in Bermùdez, Marcel, Eilan, 1995, pp. 291–309.

Carruthers, G., (2008) *Types of body representation and the sense of embodiment*, «Consciousness and Cognition», 17:1302–1316.

Cassam Q., (1995) *Introspection and bodily self-ascription*, in Bermùdez, Marcel, Eilan, 1995, pp. 311–336.

Cassam Q., (2011) *The embodied self*, in Gallagher, 2011, pp. 139-156.

Clark, A., Kiverstein, J., Tillmann, V. (a cura di), (2013) *Decomposing the will*, Oxford University Press, Oxford–New York.

De Preester, H., Knockaert, V. (a cura di), (2005) *Body image and body schema. Interdisciplinary perspectives on the body*, John Benjamins, Amsterdam–Philadelphia.

DE VIGNEMONT, F., (2007) *Habeas corpus: The sense of ownership of one's own body*, «Mind and Language», 22:427–449.

DE VIGNEMONT, F., (2010) *Body schema and body image – Pros and cons*, «Neuropsychologia», 48:669-680.

DE VIGNEMONT, F., (2011) *Bodily awareness*, in Zalta, E.N. (a cura di), «The Stanford Encyclopedia of Philosophy», (Fall 2011 Edition) http://plato.stanford.edu/archives/fall2011/entries/bodily-awareness/.

DOKIC, J., (2003) *The sense of ownership: An analogy between sensation and action*, in Roessler, Eilan, 2003, pp. 321–344.

FRITH, C.D., BLAKEMORE, S.J., WOLPERT, D.M., (2000) *Abnormalities in the awareness and control of action*, «Philosophical Transactions Royal Society of London. Series B: Biological Sciences», 355(1404):1771–1788.

FUGALI, E., (2012) *I limiti del mio corpo sono i limiti del mio mondo. Il tema del corpo proprio nella riflessione filosofica contemporanea e nella scienza cognitiva incarnata*, «Reti, Saperi, Linguaggi», 1(2):48–56.

FUGALI, E., (2016) *Fleshly matter. The constitution of the lived body: Cognitive science models and phenomenological accounts*, in Le Moli, Cicatello (a cura di), *Understanding matter. Vol. 2. Contemporary lines*, New Digital Frontiers, Palermo, pp. 65–80.

GALLAGHER, S., (2000) *Philosophical conceptions of the self: Implications for cognitive science*, «Trends in Cognitive Science», 4:14–21.

GALLAGHER, S., (2005) *How the body shapes the mind*, Oxford University Press, Oxford.

GALLAGHER, S. (a cura di), (2011), *The Oxford handbook of the self*, Oxford University Press, Oxford–New York.

GALLESE, V., SINIGAGLIA, C., (2010) *The bodily self as*

power for action, «Neuropsychologia», 48:746–755.
GALLESE, V., SINIGAGLIA, C., (2011) *How the body in action shapes the self*, «Journal of Consciousness Studies», 18:117–143.
HEAD, H., HOLMES, H.G., (1911/1912) *Sensory disturbances from cerebral lesions*, «Brain», 34:102–254.
HUSSERL, E., (1952) *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch. Phänomenologische Untersuchungen zur Konstitution*, in Id., *Gesammelte Werke*, Bd. IV, Dordrecht-Boston-London, Kluwer (trad. it. *Idee per una fenomenologia pura e per una filosofia fenomenologica. Libro secondo. Ricerche fenomenologiche sopra la costituzione*, vol. II, Einaudi, Torino, 2002).
LAKOFF, G., JOHNSON, M., (1999) *Philosophy in the flesh: The embodied mind and its challenge to western thought*, Basic Books, New York.
LE MOLI, A., CICATELLO, A., (2016) (a cura di), *Understanding matter. Vol. 2. Contemporary lines*, New Digital Frontiers, Palermo.
LONGO, M.R., AZAÑÓN, E., HAGGARD, P., (2010) *More than skin deep: Body representation beyond primary somatosensory cortex*, «Neuropsychologia», 48:655–668.
LONGO, M.R., HAGGARD, P., (2012) *What is it like to have a body*? «Current Direction in Psychological Science», 21:140–145.
LONGO, M.R., SCHÜÜR, F., KAMMERS, M.P.M., TSAKIRIS, M., HAGGARD, P., (2008) What is embodiment? A psychometric approach, «Cognition», 107:978–998.
MERLEAU–PONTY, M., (1945) *Phénoménologie de la perception*, Gallimard, Paris (trad. it. *Fenomenologia della percezione*, Bompiani, Milano, 2003).
METZINGER, T., (2010) *The self–model theory of subjectiv-*

ity: A brief summary with examples, «HumanaMente», 14:25–52.

O'SHAUGHNESSY, B., (2000) *Consciousness and the world*, Clarendon Press, Oxford.

O'SHAUGHNESSY, B., (2008) *The will: dual aspect theory*, Cambridge University Press, Cambridge.

ROESSLER, J., EILAN, N. (a cura di) (2003), *Agency and self-awareness*, Oxford University Press, Oxford-New York.

SARTRE, J.P., (1943) *L'Être et le néant: Essai d'ontologie phénoménologique*, Gallimard, Paris (trad. it., *L'essere e il nulla. Saggio di ontologia fenomenologica*, Il Saggiatore, Milano, 2002).

STAMENOV, M. L., (2005) *Body schema, body image and mirror neurons*, in De Preester, Knockaert, 2005, pp. 21–43.

SYNOFZIK, M., VOSGERAU, G., NEWEN, A., (2008) *I move, therefore I am: A new theoretical framework to investigate agency and ownership*, «Consciousness and Cognition», 17:411–424.

TSAKIRIS, M., FOTOPOULOU, A., (2013) *From the fact to the sense of agency*, in Clark, Kiverstein, Tillmann, 2013, pp. 103-117.

TSAKIRIS, M., HAGGARD, P., (2005) *Experimenting with the acting self*, «Cognitive Neuropsychology», 22:387–407.

TSAKIRIS, M., PRABHU, G., HAGGARD, P., (2006) *Having a body* versus *moving your body: How agency structures body-ownership*, «Consciousness and Cognition», 15:423–432.

TSAKIRIS, M., LONGO, M.R., HAGGARD, P., (2010) *Having a body versus moving your body: Neural signatures of agency and body-ownership*, «Neuropsychologia», 48:2740–2749.

TSAKIRIS, M., SCHÜTZ-BOSBACH, S., GALLAGHER, S., (2007) *On agency and body-ownership: Phenomenological and neurocognitive reflections*, «Consciousness and Cognition», 16:645–660.

WITTGENSTEIN, L., (1958) *The Blue and Brown Books. Preliminary studies for the «Philosophical Investigations»*, Oxford, Blackwell (trad. it. *Libro blu e libro marrone*, Einaudi, Torino, 1983).

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130667
pp. 119-139 (marzo 2018)

I processi diagnostici e le teorie dei concetti

di Francesco Gagliardi[1]

1. Introduzione

In medicina, la diagnosi è il processo che consiste nel riconoscere una condizione patologica in base ai segni clinici (oggettivi) e ai sintomi (soggettivi) del paziente. La comprensione del procedimento diagnostico è un problema affrontato con differenti finalità e metodi da diverse discipline come la filosofia, la psicologia e l'intelligenza artificiale.

La filosofia vede nel processo diagnostico un'interessante attività scientifica da comprendere e da analizzare non solo relativamente all'ambito della filosofia della medicina (Federspil, 1980; Federspil, Giaretta, 2004; Scandellari, 2004; Stempsey, 2005) ma anche più in generale per i legami che la comprensione del processo diagnostico può avere con le indagini sulla natura dei concetti (Cordeschi, Frixione 2011; Gagliardi 2014a; Murphy, 2002; Thagard, 2005) e sul ragionamento scientifico (Langley, *et al.*, 2006; Magnani, Nersessian, Thagard, 1999; Rosenblueth, Wiener, 1945; Shrager, Langley, 1990; Thagard, 1998).

[1] Independent Scholar, ORCID: 0000–0002–4270-1636. E–mail: fnc.research@gmail.com

In psicologia cognitiva il processo diagnostico è stato studiato (Cantor, *et al.*, 1980; Kihlstrom, McGlynn, 1991; Kim, Ahn, 2002; Kim, Keil, 2003) come un interessante caso particolare dei processi di categorizzazione: «*Diagnosis is an act of categorization, and as our understanding of categorization has evolved, our understanding of the diagnostic process has evolved right along with it*» (Kihlstrom riportato in Benson, 2002).

L'intelligenza artificiale ha avuto nella medicina e in particolare nella diagnosi automatica uno dei settori applicativi privilegiati[2] anche per le possibilità di indagare attraverso la simulazione dei processi diagnostici le caratteristiche del ragionamento umano: *«l'attività diagnostica in medicina è soltanto un esempio particolare di un "processo cognitivo" che sottende molte attività umane»* (Giani, 1989, p. 7).

In questo lavoro ci proponiamo di analizzare le varie tipologie di diagnosi così come analizzate e formalizzate in filosofia della medicina (le così dette *teorie della diagnosi*) alla luce delle diverse teorie dei concetti che sono state proposte nelle scienze cognitive al fine di mostrare i legami esistenti tra le teorie della diagnosi e le teorie dei processi di categorizzazione.

Nel seguito introduciamo preliminarmente le principali teorie dei concetti mostrandone le caratteristiche essenziali, quindi illustriamo gli elementi fondamentali per comprendere il processo diagnostico e le varie teorie proposte per spiegare i vari tipi di diagnosi. Illustreremo quindi come sia possibile legare le varie teorie della diagnosi con le teorie dei concetti, mostrando anche i possibili legami con

[2] Si pensi ai primi sistemi esperti come MYCIN (SHORTLIFFE, 1976), realizzato per diagnosticare infezioni del sangue, fino all'attuale sviluppo dei *Clinical Decision Support System* (COIERA, 2003, Cap. 25) e del *medical data mining* (CIOS, MOORE, 2002).

alcuni risultati di psicologia cognitiva, intelligenza artificiale e filosofia della scienza.

2. Categorizzazioni e teorie dei concetti

La categorizzazione è un processo cognitivo fondamentale grazie al quale gli esseri umani "ritagliano" e danno un senso alla realtà che li circonda (Gagliardi, 2014a; Houdé, 1998; Kruschkea, 2001; Medin, Aguilar, 1999); attraverso i processi di categorizzazione la mente umana suddivide il mondo in categorie costruendo dei concetti che forniscono le rappresentazioni mentali di queste categorie.

I concetti sono stati definiti come una sorta di *«colla mentale»* (Murphy 2002, p. 1), che lega le esperienze passate con le attuali interazioni col mondo e sono le *«forme di conoscenza parziale e "prospettica" con cui gli umani, creature finite e compromesse, cercano di ordinare e di "dare un senso" alla realtà che li circonda»* (Gagliardi 2014a, p. 294).

Le principali teorie sulla natura dei concetti (Medin, 1989; Murphy, 2002; Thagard, 2005) che illustriamo nel seguito sono: la teoria classica, la teoria dei prototipi, la teoria degli esemplari, e la *theory–theory*.

2.1 *La teoria classica*

Secondo la teoria classica (Frege 1891; Frege 1892; Zalta, 2010) un concetto è definito da alcune caratteristiche che sono le condizioni necessarie e sufficienti per la sua definizione e sono espresse per mezzo di predicati logici. Ogni oggetto che soddisfa il predicato logico appartiene alla categoria rappresentata da quel predicato. Secondo questa

teoria, un oggetto appartiene ad una data categorie in modo netto e senza ambiguità, ovvero secondo la regola del terzo escluso: o appartiene o non appartiene ad una categoria. Inoltre, qualsiasi oggetto soddisfi la definizione è un membro a pieno titolo della categoria come qualsiasi altro oggetto soddisfi le condizioni, ovvero non c'è "gradazione" nell'appartenenza ad una categoria.

2.2 *La teoria dei prototipi*

La prima teoria proposta per superare alcuni dei problemi della teoria classica della categorizzazione è la teoria dei prototipi (Rosch, 1975; Rosch, 1978; Rosch, Mervis, 1975) (si veda anche l'idea di "somiglianza di famiglia" di Wittgenstein [1974, pp. 46-47]); secondo questa teoria i concetti sono dei prototipi che rappresentano le caratteristiche tipiche degli oggetti di una categoria piuttosto che le condizioni necessarie e sufficienti. Secondo la teoria dei prototipi gli esseri umani tendono a identificare una categoria di oggetti e a ragionare a proposito dei propri membri, facendo riferimento ad un oggetto preciso tipico della famiglia.

2.3 *La teoria degli esemplari*

Secondo la teoria degli esemplari (Medin, Schaffer, 1978) i concetti sono una collezione di esempi memorizzati. Essa è radicalmente diversa dalle precedenti teorie poiché rigetta l'idea, comune alla teoria classica e a quella dei prototipi, che le persone abbiano un qualche tipo di rappresentazione capace di descrivere l'intera categoria.

2.4 *La theory–theory*

La teoria della teoria (Gopnik, Meltzoff, 1997; Medin 1989; Murphy, Medin 1985) considera i concetti come parte della nostra conoscenza generale del mondo, e non solo rispetto al tipo di rappresentazione delle categorie (predicati logici, prototipi, esemplari).

In accordo con questa teoria i concetti hanno un'essenziale funzione esplicativa piuttosto che descrittiva, costituendo una sorta di modello del mondo osservato. Questa teoria non è alternativa a quelle precedenti, e si può considerare costruita sopra di esse (Murphy, 2002, p. 60), infatti la teoria della teoria è compatibile con l'idea che i concetti siano descrizioni di qualche tipo, come ad esempio i prototipi o gli esemplari (cfr. Gagliardi, 2016).

3. Processi diagnostici e teorie della diagnosi

Nell'ambito medico il termine "diagnosi" ha almeno tre significati, potendosi riferire alla malattia di cui è affetto un paziente, al ragionamento clinico e alla conoscenza clinica utilizzata nello stesso (si veda Giani, 1989, pp. 32–34)[3].

Per quanto riguarda il primo significato, il termine "diagnosi" viene usato per denotare la malattia di cui è affetto un paziente, come ad esempio nella seguente frase: «*La diagnosi del sig. Rossi è psoriasi*». In questa accezione il–significato di "diagnosi" non è connotato ontologicamente poiché non si sta affermando che «*La malattia del sig. Rossi è psoriasi*». In quest'ultima frase si afferma

[3] Nella successiva descrizione di questi tre significati seguiamo in parte lo stesso autore.

qualcosa di "assoluto" sullo stato di salute del paziente, e ciò non è rispondente alla reale attività clinica della diagnosi. Infatti, nella pratica medica si effettuano diagnosi congetturali[4] e diagnosi "rivedibili", come ad esempio nelle procedure ospedaliere in cui è prevista una "diagnosi di ingresso" e una "diagnosi di dimissione", o anche nei casi di errori diagnostici (cfr. Skrabanek, McCormick, 1995, p. 81). Il termine "diagnosi" in questa accezione ha quindi una valenza epistemica piuttosto che ontologica e denota solo la malattia che viene individuata al termine di un ragionamento clinico.

Tabella 1. Tre significati della diagnosi

	Diagnosi	**Significati**
D_1	Diagnosi_1	La malattia diagnosticata
D_2	Diagnosi_2	Il ragionamento diagnostico
D_3	Diagnosi_3	La conoscenza utilizzata nel ragionamento diagnostico

Il secondo significato del termine "diagnosi" si riferisce al ragionamento diagnostico cioè al modo attraverso il quale il diagnosta, dato un particolare caso clinico, individua una malattia ovvero giunge alla "diagnosi" nel significato precedente; ad esempio nella seguente frase: «*La diagnosi del sig. Rossi è stata lunga e difficoltosa*» ci si riferisce alla difficoltà incontrata nel ragionamento che ha portato a individuare la psoriasi.

Il terzo significato del termine "diagnosi" è relativo alla conoscenza scientifica e al sapere medico usati dal diagno-

[4] Sulla distinzione tra diagnosi categoriale e diagnosi congetturale si veda (SADEGH–ZADEH, 2000, p.229).

sta per effettuare le diagnosi, come ad esempio nella seguente frase: «*La diagnosi di psoriasi si basa sulla presenza di eritema, acantosi, infiltrato mononucleare*»*;* in questa accezione la diagnosi ha un carattere in parte normativo e riguarda sia la conoscenza medica condivisa, come nel precedente esempio, che la personale esperienza clinica del diagnosta.

Il termine diagnosi ha dunque tre diversi significati, «*spesso non messi chiaramente in luce nei testi di medicina*» (Giani, 1989, p. 34), che nel seguito, ove necessario, indichiamo rispettivamente come $Diagnosi_1$, $Diagnosi_2$ e $Diagnosi_3$ (v. tabella 1).

In sintesi la diagnosi intesa come processo diagnostico di tipo epistemico–cognitivo si può formalizzare come una funzione che data la conoscenza scientifica utilizzata del medico e considerato un particolare paziente, che indichiamo con P, ha come risultato la diagnosi intesa come malattia, ovvero usando la simbologia introdotta prima si ha:

$$D_2\left(P, D_3\right) \rightarrow D_1 \qquad (1)$$

dove D_1, è la malattia diagnosticata, con i suoi limiti epistemologici indicati prima, D_2 è il ragionamento diagnostico effettuato, ovvero il particolare processo cognitivo eseguito e D_3 è la conoscenza impiegata nel ragionamento diagnostico.

In filosofia della medicina sono state introdotte le cosiddette *teorie della diagnosi* con cui si cerca di comprendere e formalizzare le varie tipologie dei processi diagnostici; le principali teorie delle diagnosi pertinenti alla diagnosi categoriale qui considerate sono la *diagnosi fisiopa-*

tologica e la *diagnosi nosologica*[5] (Gioriello, Moriggi, 2004, p. 10; Sadegh–Zadeh, 2000, p. 230; Scandellari, 1981, pp. 55–56):

a) La diagnosi fisiopatologica detta anche *diagnosi causale* è la prassi diagnostica in cui si procede alla spiegazione delle cause dei fenomeni morbosi riscontrati nel paziente utilizzando le conoscenze della fisiologia umana. La diagnosi è dunque ottenuta legando questa ai dati clinici attraverso la ricostruzione di un nesso causale.
b) La diagnosi nosologica è la prassi diagnostica in cui si presta più attenzione all'insorgenza di complessi sindromici "tipici". La diagnosi è ottenuta analizzando la similarità del singolo caso clinico con i vari quadri morbosi con cui è noto che si manifestino le patologie.

Queste due teorie si differenziano tra loro per il diverso ragionamento diagnostico (*diagnosi*$_2$) e la diversa conoscenza scientifica su cui si basano (*diagnosi*$_3$); nel seguito analizziamo separatamente le due teorie della diagnosi mettendole in relazione alle teorie dei processi di categorizzazione.

4. Le teorie della diagnosi e le teorie dei concetti

La diagnosi fisiopatologica si basa, in un certo senso, sulla costruzione di una teoria che spieghi il caso clinico consi-

[5] Alcuni autori (e.g. SADEGH-ZADEH, 2000, p. 230) introducono un terzo tipo di diagnosi, la diagnosi di anormalità di cui la diagnosi nosologica è un tipo particolare.

derato in relazione ad un modello della fisiologia umana, è dunque come mostriamo nel seguente paragrafo una categorizzazione riconducibile principalmente alla *theory–theory*; mentre la diagnosi nosologica si basa sulla nozione di similarità presentando aspetti riconducibili sia alla teoria dei prototipi, sia a quella degli esemplari (v. tabella 2).

4.1. *La diagnosi fisiopatologica e la teoria della teoria*

La diagnosi causale si può considerare un processo di categorizzazione in cui dato un paziente *P* e un modello della normale fisiologia umana, che indichiamo con *M*, si perviene alla diagnosi:

$$D_{causale}\left(P, M\right) \rightarrow D_1 \qquad (2)$$

La diagnosi fisiopatologica o causale è dunque un processo di ricostruzione *a posteriori* dei nessi causali che hanno determinato l'insorgere dei sintomi e segni clinici rispetto ad un normale stato fisiologico. Questo tipo di processo diagnostico consiste nell'elaborazione di una congettura plausibile o di una teoria che spieghi il caso clinico in esame e quindi lo possiamo considerare un processo di categorizzazione *theory–based*.

A supporto di tale tesi, si può osservare che alcuni lavori di psicologia sperimentale, che hanno considerato come soggetti di studio proprio dei medici durante la propria attività clinica, hanno messo in evidenza come la diagnosi sia un processo di categorizzazione affetto anche dalle proprie congetture e conoscenze mediche (de Kwaadsteniet, Kim, Yopchick, 2013; Kim, Ahn, 2002; Kim, Keil, 2003) oltre che dal corretto inquadramento in una tassonomia nosologica (v. dopo) (cfr. Sadegh–Zadeh, 1999).

Anche in intelligenza artificiale, alcuni approcci computazionali della formalizzazione dei processi diagnostici, non solo in ambito medico, sono basati sulla disponibilità di un modello del funzionamento del sistema considerato di cui si vuole effettuare una diagnosi per individuare le cause del "guasto" manifestatosi (Console, Torasso, 2006; cfr. Langley, *et al.*, 2006).

Inoltre, questo tipo di diagnosi è analogo ad altre forme di ragionamento scientifico basate sulla costruzione di modelli e teorie, ovvero a quello che in filosofia della scienza si considera come il *model based reasoning,* che è adottato da molte discipline scientifiche e non solo della medicina (Magnani, Nersessian, Thagard, 1999; Rosenblueth, Wiener, 1945; Shrager, Langley, 1990; Thagard, 1998).

4.2. *La diagnosi nosologica, i prototipi e gli esemplari*

La diagnosi nosologica è un processo di categorizzazione in cui dato un paziente si cerca di "riconoscere" la malattia di cui è affetto, considerando i segni clinici e i sintomi che manifesta senza dare necessariamente un'interpretazione o spiegazione del quadro morboso[6].

La diagnosi nosologica è un processo di categorizzazione basato sulla similarità che si può mettere in relazione sia alla teoria dei prototipi che a quella degli esemplari avendo aspetti riconducibili ad entrambi le teorie.

---

[6] Augusto Murri, noto come "sommo dei clinici medici" la cui fama come diagnosta rivaleggiava con quella di Antonio Cardarelli sentenziava: *«in clinica si deve soprattutto riconoscere»* (MURRI, 1905/1972, p. 11) poiché *«l'analisi sperimentale qui* [in clinica] *è concessa di rado»* (MURRI, 1905/1972, p. 14) (per un commento si veda GIORIELLO, MORIGGI, 2004, pp. 9-10).

La sindrome è definita (Ford–Martin, 2002, p. 1496; Gagliardi, 2014b) come un insieme di caratteristiche clinicamente riconoscibili che spesso tendono a presentarsi insieme in forma simile e che può essere la manifestazione di una particolare patologia; la sindrome è quindi un prototipo di un insieme di osservazioni cliniche, un quadro morboso tipico di una data patologia, che si manifesta in maniera simile in diversi casi.

La diagnosi basata sul riconoscimento di una sindrome, ovvero sulla similarità tra il paziente considerato e il quadro morboso sindromico, che è tipico di una patologia, è da considerare un tipo di diagnosi in accordo con la teoria dei prototipi.

I processi diagnostici basati sulla similarità non coinvolgono solo le sindromi, ma possono basarsi anche su singoli casi clinici precedentemente noti al diagnosta. Infatti sin dalle prime fasi di formazione e specializzazione i medici sono incoraggiati a considerare ogni paziente come un potenziale caso di studio utile per le future attività diagnostiche (Wyngaarden, 1979)[7].

La diagnosi basata sulla similarità tra un precedente caso clinico, anche atipico, di cui sia già nota la diagnosi e il paziente considerato sono una forma di categorizzazione basata sugli esemplari.

Unificando entrambi questi aspetti, la diagnosi nosologica è un processo di categorizzazione in cui dato un paziente P, e un insieme di quadri morbosi Q, di cui sia nota la malattia (che possono essere sia delle sindromi tipiche e

[7] Si veda anche (FEATHERSTONE, BEITMAN, IRBY, 1984; SKRABANEK, MCCORMICK, 1995, p. 73) per i possibili effetti negativi di una tale forma di apprendimento; si veda anche (GAGLIARDI, 2011; GIARETTA, 2004) sull'intreccio tra aspetti generali (nomotetici) e particolari (idiografici) nel processo diagnostico.

ben note, che dei singoli casi clinici noti al diagnosta) si giunge alla diagnosi:

$$D_{Nosologica}\left(P,Q\right)\rightarrow D_1 \tag{3}$$

La diagnosi nosologica, intesa come diagnosi basata sulla similarità, è dunque un processo di riconoscimento della malattia che possiamo considerare come un processo di categorizzazione misto prototipi ed esemplari che non può essere ricondotto esclusivamente ad una sola di queste teorie dei concetti[8].

Anche per la diagnosi nosologica ci sono precedenti lavori in psicologia sperimentale, filosofia della scienza e intelligenza artificiale che sono legati a questo tipo di diagnosi, ma a differenza del caso della diagnosi fisiopatologica, queste discipline si sono occupate principalmente solo di aspetti parziali di essa.

L'intelligenza artificiale si è interessata principalmente al *case–based reasoning* (basato su esemplari; Bichindaritz, 2006; Voskoglou, 2008) anche riguardo ai suoi legami con la filosofia della scienza (Aha, 1998) e solo secondariamente ai sistemi basati su prototipi, (e.g. Everitt, *et al.*, 1971); d'altra parte la psicologia cognitiva sperimentale si è occupata principalmente della diagnosi nosologica basata solo su sindromi prototipiche (Cantor, *et al.*, 1980; Livesley, 1991).

Solo in tempi relativamente recenti sono stati proposti nell'ambito multidisciplinare della scienza cognitiva e dell'intelligenza artificiale in medicina dei sistemi di diagnosi automatica capaci di compiere una

[8] Sull'infondatezza metodologica della contrapposizione tra teoria dei prototipi e teoria degli esemplari si veda (GAGLIARDI, 2009).

diagnosi nosologica basata sia sulle sindromi che sui casi clinici atipici (Gagliardi, 2011, 2013); questi sistemi sono realizzati con un modello computazionale della categorizzazione, il *PEL–C* (*Prototype Exemplar Learning Classifier)*, capace di combinare insieme la categorizzazione basata su prototipi e quella basata su esemplari (Gagliardi, 2008, 2012).

Tabella 2. Quadro sinottico dei legami tra il ragionamento diagnostico ($diagnosi_2$), la conoscenza clinica utilizzata ($diagnosi_3$) e le teorie dei concetti.

$Diagnosi_2$	**$Diagnosi_3$**	**Teorie dei Concetti**
Diagnosi Nosologica	Sindromi	T. dei Prototipi
	Casi clinici	T. degli Esemplari
Diagnosi Causale	Fisiologia	T. della Teoria

5. Conclusioni

La diagnosi è un'attività che riguarda tre aspetti clinici diversi: la malattia di cui è affetto un paziente, il ragionamento attraverso il quale il diagnosta giunge a individuare una malattia e la conoscenza scientifica usata a tal fine.

Il processo diagnostico viene compiuto, secondo le teorie della diagnosi, attraverso due principali modalità: la diagnosi fisiopatologica e quella nosologica; entrambi sono da considerarsi dei processi cognitivi di categorizzazione di una data condizione morbosa.

In questo lavoro abbiamo mostrato che entrambi si possono considerare come dei particolari processi cognitivi di

categorizzazione e concettualizzazione della mente umana effettuati nell'ambito clinico; la diagnosi fisiopatologica è una forma di ragionamento causale *model–based* che rientra nella *theory–theory* dei processi di categorizzazione; la diagnosi nosologica, invece, basandosi sulla similarità è un'attività di categorizzazione con aspetti riconducibili sia alla teoria dei prototipi quanto alla teoria degli esemplari.

Abbiamo inoltre legato questa nostra analisi con alcuni risultati della psicologia cognitiva, dell'intelligenza artificiale e della filosofia della scienza.

Le due teorie della diagnosi non sono in contrasto tra di loro così come non devono considerarsi in contrasto le teorie della categorizzazione basate su prototipi, su esemplari e sulle teorie; lo studio dei processi diagnostici e di categorizzazione, piuttosto che tendere a contrappore le varie teorie esistenti, dovrebbe mirare a comprendere le capacità di integrazione e uso, da parte della mente umana, dei vari tipi di ragionamento e di categorizzazione.

Riconoscimenti

Alcuni aspetti preliminari di questo lavoro sono apparsi in Gagliardi (2010).

Riferimenti bibliografici

AHA, D.W. (1998) *The omnipresence of case–based reasoning in science and application*, «Knowledge–Based Systems», 11(5/6):261–273.

BENSON, E. (2002) *Thinking clinically. A new study shows how clinicians' theories could affect their diagnoses*.

«APA Monitor on Psychology», 33(11):30.

BICHINDARITZ, I (2006) *Case–based reasoning in the health sciences (Guest Editorial)*, «Artificial Intelligence in Medicine», 36(2):121–125.

CANTOR, N., SMITH, E.E., FRENCH, R., MEZZICH, J. (1980) *Psychiatric diagnosis as prototype categorization*, «Journal of Abnormal Psychology», 89(2):181–193.

CIOS KJ, MOORE GW. (2002) *Uniqueness of medical data mining. «Artificial Intelligence in Medicine»*, 26(1/2):1–24.

COIERA, E. (2003) *Guide to Health Informatics, 2nd ed*., Hodder & Stoughton. ISBN:0340764252.

CONSOLE, L., TORASSO, P. (2006) *Automated Diagnosis*, «Intelligenza Artificiale». 3(1/2):42–48.

CORDESCHI, R., FRIXIONE, M. (2011) *Rappresentare i concetti: filosofia, psicologia e modelli computazionali*, «Sistemi Intelligenti» 23(1):25–40.

DE KWAADSTENIET, L., KIM, N.S., YOPCHICK, J.E. (2013) *How do practising clinicians and students apply newly learned causal information about mental disorders?* «Journal of Evaluation in Clinical Practice», 19(1):112–117.

EVERITT, B.S., GOURLAY, A.J., KENDELL R.E. (1971) *An attempt at validation of traditional psychiatric syndromes by cluster analysis*, «British Journal of Psychiatry». 119(551):399–412.

FEATHERSTONE, H.J., BEITMAN, B.D., IRBY, D.M. (1984) *Distorted learning from unusual medical anecdotes*, «Medical Education», 18(3):155–158.

FEDERSPIL, G. (1980) *I fondamenti del metodo in medicina clinica e sperimentale*, Piccin editore, Padova.

FEDERSPIL, G., GIARETTA, P. (2004) (a cura di) *Forma della Razionalità Medica*, Rubbettino Scientifica, Catanzaro.

FORD–MARTIN, P.A. (2002) *Key terms. Syndrome*, in Longe, J.L., Blanchfield, D.S. (eds.) *Gale Encyclopedia of Medicine, 2nd edn.,* p. 1496, Gale Group, Farmington Hills, MI.

FREGE, G. (1891) *Funktion und Begriff*. Verlag Hermann Pohl, Jena, Germany. (traduzione italiana di Picardi E. (2001) *Funzione e concetto*, in Penco, C. Picardi, E. (2001) (a cura di) *Frege: Senso, funzione e concetto. Scritti filosofici*, Laterza editore, Roma–Bari.

FREGE, G. (1892) *Über Begriff und Gegenstand*, «Philosophie», XVI, 192–205. (traduzione italiana di Zecchi, S. (1973) *Concetto e oggetto*, in Bonomi, A. (a cura di) *La struttura logica del linguaggio*, Bompiani editore, Milano, Italy.

GAGLIARDI, F. (2008) *A Prototype–Exemplars Hybrid Cognitive Model of "Phenomenon of Typicality" in Categorization: A Case Study in Biological Classification,* in Love, B.C., McRae, K., Sloutsky, V.M. (eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, pp. 1176–1181.

GAGLIARDI, F. (2009) *La categorizzazione tra psicologia cognitiva e machine learning: perché è necessario un approccio interdisciplinare,* «Sistemi Intelligenti», 21(3):489–501.

GAGLIARDI, F. (2010) *Teorie della Diagnosi e Teorie della Categorizzazione*, in Ferrari, G., Bouquet, P., Cruciani, M., Giardini, F. (a cura di) *Pratiche della Cognizione. Atti del Settimo Convegno Nazionale di Scienze Cognitive*. Università degli Studi di Trento Editore, Trento, pp. 213–217.

GAGLIARDI, F. (2011) *Instance–based classifiers applied to medical databases: diagnosis and knowledge extraction*, «Artificial Intelligence in Medicine», 52(3):123–

139.

GAGLIARDI, F. (2012) *Modelling Typicality in Categorization with Instance–Based Machine Learning*, «Cognitive Systems», 7(3):275–293.

GAGLIARDI, F. (2013) *Le applicazioni delle Teorie della Categorizzazione ai sistemi per la diagnosi automatica (Abstract della Relazione ad Invito)*, in Cruciani, M. (a cura di) *Le scienze cognitive: applicazioni e valore socio–economico*, Università degli Studi di Trento Editore, Trento. p. 21.

GAGLIARDI, F. (2014a) *La naturalizzazione dei concetti: aspetti computazionali e cognitivi*, «Sistemi Intelligenti», 26(2):283–295.

GAGLIARDI, F. (2014b) *A Cognitive Machine–Learning System to Discover Syndromes in Erythemato–Squamous Diseases*, in Rodrigues, J. (ed.) *Advancing Medical Practice through Technology: Applications for Healthcare Delivery, Management, and Quality*, IGI Global Publisher, Hershey, PA, pp. 66–101.

GAGLIARDI, F. (2016) *La concettualizzazione dell'antimateria tra permeabilità cognitiva, categorizzazione embodied e theory–based*, «Sistemi Intelligenti», 28(1):105–124.

GIANI, U. (1989) *La mente diagnostica. Probabilità, incertezza e modelli di Intelligenza Artificiale in Medicina*. Liguori Editore, Napoli.

GIARETTA P (2004) *Aspetti idiografici e nometetici del procedimento clinico: analisi di un caso*. In Federspil G., Giaretta P. (2004) (a cura di) *Forma della Razionalità Medica*, Rubbettino Scientifica, Catanzaro, pp. 143–162.

GIORIELLO, G., MORIGGI, S. (2004) *Tra diagnosi e scoperta. Una rilettura del caso Semmelweis*, in Federspil G., Giaretta P. (2004) (a cura di) *Forme della Razionalità*

Medica, Rubbettino Scientifica, Catanzaro, pp. 9–30.

GOPNIK, A., MELTZOFF, A. (1997) *Words, Thoughts, and Theories*, MIT Press, Cambridge, MA.

HESSLOW G. (1993) *Do we need a concept of disease?* «Theoretical Medicine». 14(1):1–14.

HOUDE, O. (1998) *Categorization*, in Houde, O., Kayser, D., Koenig, O., Proust, J., Rastier, F. (eds.) *Vocabulaire de sciences cognitives*, Presses Universitaires de France, Paris. (Traduzione italiana: Houde, O. *et al*. (2000) *Dizionario di scienze cognitive. Neuroscienze, psicologia, intelligenza artificiale, linguistica, filosofia*, Editori Riuniti, Roma.

KIHLSTROM, J.F., MCGLYNN, S.M. (1991) *Experimental research in clinical psychology*, in Hersen, M., Kazdin, A. E., Bellack, A. S. (eds.) *Clinical psychology handbook, 2nd edn.*, Pergamon. New York, NY, pp. 239–257.

KIM, N.S., AHN, W.K. (2002) *Clinical psychologists' theory–based representations of mental disorders predict their diagnostic reasoning and memory*, «Journal of Experimental Psychology: General», 131(4):451–476.

KIM, N.S., KEIL, F.C. (2003) *From symptoms to causes: Diversity effects in diagnostic reasoning*, «Memory & Cognition», 31(1):155–165.

KRUSCHKEA, J.K. (2001) *Categorization and Similarity Models*, in Smelser, N.J., Baltes, P.B. (eds.) *International Encyclopedia of the Social & Behavioral Sciences*, Pergamon Press, Oxford, UK, pp. 1532–1535.

LANGLEY, P., SHIRAN, O., SHRAGER, J., TODOROVSKI, L. POHORILLE, A. (2006) *Constructing explanatory process models from biological data and knowledge*, «Artificial Intelligence in Medicine», 37(3):191–201.

LIVESLEY, W. J. (1991) *Classifying personality disorders: Ideal types, prototypes, or dimensions?* «Journal of

Personality Disorders», 5(1), pp. 52–59.

MAGNANI, L., NERSESSIAN, N.J., THAGARD, P. (1999) (eds.) *Model–Based Reasoning in Scientific Discovery*, Kluwer Academic Publishers/Plenum Publishers, New York, NY.

MEDIN, D.L. (1989) *Concepts and conceptual structure*, «*American Psychologist*», 44(12):1469–1481.

MEDIN, D.L., AGUILAR, C. (1999) *Categorization*, in Wilson, R.A., Keil, F. (eds.) *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, MIT Press, Cambridge, MA, pp. 104–106.

MEDIN, D.L., SCHAFFER, M.M. (1978) *Context theory of classification learning,* «Psychological Review», 85(3):207–238.

MURPHY, G.L. (2002) *The big book of concepts, MIT Press, Cambridge*, MA.

MURPHY, G.L., MEDIN, D.L. (1985) *The role of theories in conceptual coherence*, «Psychological Review», 92(3):289–316.

MURRI A. (1972) *Quattro lezioni e una perizia, Il problema del metodo in medicina e biologia*, Zanichelli, Bologna. (Edizione originale del 1905).

ROSCH, E. (1975) *Cognitive Representations of Semantic Categories*. «Journal of Experimental Psychology», 104(3):192–233.

ROSCH, E. (1978) *Principles of Categorization*, in Rosch, E., Lloyd, B. B. (eds.) *Cognition and Categorization*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 27–48.

ROSCH, E., MERVIS, C. (1975) *Family Resemblances: Studies in the Internal Structure of Categories*, «Cognitive Psychology», 7(4):573–605.

ROSENBLUETH, A., WIENER, N. (1945) *The role of Models in Sciences*. «Philosophy of Science». 12(4):316–321.

SADEGH–ZADEH, K. (1999) *Fundamentals of clinical methodology: 3. Nosology*, «Artificial Intelligence in Medicine», 17(1):87–108.

SADEGH–ZADEH, K. (2000) *Fundamentals of clinical methodology: 4. Diagnosis*. «Artificial Intelligence in Medicine», 20(3):227–241.

SCANDELLARI, C. (1981) *La strategia della diagnosi*. Piccin editore, Padova.

SCANDELLARI, C. (2004) *La diagnosi clinica. Principi metodologici del procedimento decisionale*. Elsevier, Collana biblioteca medica Masson, Milano.

SHORTLIFFE, E.H. (1976) *Computer–Based Medical Consultations: MYCIN*. Elsevier North–Holland, Amsterdam, London, New York.

SHRAGER, J., LANGLEY, P. (1990) (eds.) *Computational Models of Scientific Discovery and Theory Formation*, San Francisco: Morgan Kaufmann.

SKRABANEK, P., MCCORMICK, J. (1995) *Follie e inganni della medicina, 2° ed.*, Marsilio Editori, Venezia.

STEMPSEY, W.E. (2005) *The philosophy of medicine: Development of a discipline*, «Medicine, Health Care and Philosophy», 7(3):243–251.

THAGARD, P. (1998) *Explaining disease: Correlations, causes, and mechanisms*, «Minds and Machines», 8(1):61–78.

THAGARD, P. (2005) *Mind: Introduction to cognitive science, 2nd edn.*, MIT Press, Cambridge, MA.

Voskoglou M.G. (2008) *Case–Based Reasoning: A Recent Theory for Problem–Solving and Learning*, in Lytras M.D., Carroll, J.M., Damiani, E., Tennyson, R.D., Avison, D., Vossen, G., De Pablos, P.O., (eds) *The Open*

Knowlege Society. A Computer Science and Information Systems Manifesto. *WSKS 2008. Communications in Computer and Information Science, vol 19*, Springer, Berlin, Heidelberg.

WITTGENSTEIN, L. (1974) *Ricerche filosofiche*, Einaudi, Torino (Edizione originale del 1953/1958, *Philosophische Untersuchungen*, *Philosophical investigations*, Blackwell, Oxford).

WYNGAARDEN, J.B. (1979) *The Clinical Investigator as an Endangered Species,* «The New England Journal of Medicine», 301(23):1254–1259.

ZALTA, E.N. (2010) *Frege's Logic, Theorem, and Foundations for Arithmetic*, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy. Fall 2010 Edition*, Stanford University, Stanford, CA. http://plato.stanford.edu/archives/fall2010/entries/frege-logic/

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130668
pp. 141-169 (marzo 2018)

Osservazioni sui processi di categorizzazione tra concettualismo e non–concettualismo

di Giuliana Gerace[1]

1. Introduzione

Un processo di categorizzazione è un processo cognitivo attraverso cui una mente organizza e fissa i dati del reale all'interno di strutture di significato relativamente stabili. La nozione di categorizzazione è stata prevalentemente indagata dalla psicologia nel suo aspetto di strumento cognitivo, al fine di comprendere il modo con cui una struttura di significato o "concetto" possa stabilire una continuità tra le esperienze.

In passato il dibattito teorico sui possibili requisiti di una struttura di significato, o concetto, in grado di rappresentare una categoria su tutte le altre ha coinvolto non solo la psicologia cognitiva ma, in parte, anche la filosofia e ha variamente considerato sia la necessità che tale struttura codifichi tutte le proprietà possedute dai membri di una categoria, sia la necessità che siano codificate solo le proprietà che marcano un criterio di identità (prototipo) in base al quale connettere le informazioni ricevute.

Recentemente tale dibattito si è spostato sulla plausibile ipotesi che una simile struttura di significato si basi sulla

[1] Ricercatore Indipendente. ORCID: 0000–0003–1715–5650. E–mail: giuliana.gerace@gmail.com

simulazione mnemonica di situazioni esperite e che, come tale, consista in una rappresentazione prospettica di diversi nessi di significato, in grado di scomporsi e ricomporsi in funzione dei contesti d'esperienza che la attivano. In questa prospettiva emerge dunque il problema di definire una cosiddetta "rappresentazione percettiva", quale nozione piuttosto distante dalla definizione tradizionale di concetto.

Nelle seguenti riflessioni, la descrizione di un'efficace configurazione delle strutture rappresentative e dei processi cognitivi che ci consentono di discriminare tra più oggetti di esperienza s'intreccia con la questione, più propriamente teorica, relativa alla possibilità di una rappresentabilità logico-semantica di queste strutture ovvero di una loro possibile formalizzazione, preludio a qualunque possibilità di ricostruzione artificiale delle stesse. A questo proposito, nell'ultima parte di queste riflessioni, saranno considerati argomenti di natura filosofica relativi al contenuto rappresentativo e, in particolare relativi al contenuto non–concettuale ovvero irriflesso della percezione.

2. Il problema di una categorizzazione stabile e flessi bile

La maggior parte degli studi legati alle nozioni di "concetto" e di "categorizzazione" si concentrano sui requisiti che le caratterizzano rispettivamente come rappresentazioni mentali e come processi cognitivi che a queste conducono (Goldstone, Kersten 2003). Le maggiori teorie che si sono susseguite in questo ambito hanno come obbiettivo la definizione di criteri, che siano per un verso stabili e "cognitivamente economici" (Collins, Quillian 1969); per altro

verso flessibili e dunque in grado di garantire la possibile varietà della conoscenza.

Come noto (Pezzulo 2007), la psicologia cognitiva considera pressoché obsoleto il processo di categorizzazione classico o aristotelico, la cui organizzazione gerarchica e inclusiva degli elementi si basa su criteri di appartenenza definitori, che si rifanno ad un insieme di proprietà prestabilite, necessarie e sufficienti[2] (Murphy 2002). Il motivo dell'inadeguatezza di questo modello è legato alla sua prospettiva rigida e altamente riduzionistica: la categorizzazione di ogni elemento individuale è riconducibile ad un insieme chiuso di proprietà invarianti, motivo per cui il modello classico rimane poco fedele alla complessa varietà dei possibili oggetti di conoscenza (Keane, Eysenck 2005).

Per contro, la diffusissima teoria dei prototipi (Rosch 1973, 1978, 1983; Rosch, Mervis 1975), si focalizza sul carattere anti–definitorio delle rappresentazioni categoriali e consente di spiegare plausibili meccanismi cognitivi, che la teoria classica non è in grado di giustificare. Se è vero che le categorie sono rappresentazioni concettuali dalle caratteristiche tendenzialmente stabili (Smith, Minda 2002), è anche vero che tali concetti non sono sempre definibili come insiemi chiusi di proprietà esplicite e sempre valide, ma spesso si presentano come costrutti integrati e covarianti, le cui caratteristiche emergono sullo sfondo di relazioni di somiglianza e/o pertinenza semantica tra le diverse proprietà che presentano (Hampton, 1979, 1993), come del resto ampiamente argomentato sul piano teorico–filosofico (Wittgenstein 1953). Al fine di identificare una categoria, sarà cruciale dunque non una lista di attributi

[2] Nella Logica aristotelica le categorie coincidono con delle qualità o "predicamenti" delle sostanze prime.

necessari e sufficienti, bensì la rappresentazione mentale o il "concetto" di un elemento in grado di esprimere i cosiddetti effetti di tipicità (Hampton, 1993) ovvero il massimo numero di tratti comuni ad altri potenziali membri di quella categoria. Tali tratti comuni possono essere anche correlazioni di vari attributi, che dunque, a differenza che nel modello definitorio classico, possono co–variare. La condizione di tipicità in termini di somiglianza tra attributi maggiormente interrelati e frequenti (che non devono necessariamente essere comuni a tutti gli esemplari) consente di rappresentare la semantica prototipale come una combinazione di insiemi non rigidi ma sfumati, in grado di configurare i diversi gradi di appartenenza ad una categoria (Alxatib, Pelletier 2011).

I vantaggi di questa prospettiva riguardano, per un verso, la stabilità "cognitivamente economica" dei processi di categorizzazione implicati (Feldman *et al.* 2009; Griffiths, Tenenbaum 2009): una rappresentazione prototipica è in grado di "captare" i gradi di similarità di un ampio numero di esemplari (Rosch 1973); per altro verso il potenziale di varietà conoscitiva derivante dall'esperienza: l'individuazione di elementi di identità rappresentativa avviene o per astrazione delle caratteristiche più comuni tra più individui o tramite la memoria di un "buon esemplare", generalmente il primo storicamente appreso nell'ambito di quella categoria o insieme (Rosch, Mervis 1975).

Nel complesso, la teoria dei prototipi, supportata da evidenze sperimentali sul giudizio di tipicità (Rosch, Mervis 1975; Hampton 1995), che confermano la valenza dei processi di categorizzazione coinvolti, anche se per lo più in termini euristici, presenta una serie di vantaggi esplicativi in termini di flessibilità semantica delle rappresentazioni. In aggiunta, la possibilità di formalizzare tale flessi-

bilità attraverso la costruzione di insiemi vaghi, rende questi processi di categorizzazione matematicamente rappresentabili (Alxatib, Pelletier 2011).

Tuttavia la valenza esplicativa di tali rappresentazioni è inefficace, se si considera che la teoria dei prototipi rimane generica con riguardo alla specificazione dei vincoli di salienza semantica ovvero di quei criteri che condizionano l'evidenza di determinati tratti primari tipici (Hampton 1993). Se si tratta solo di vincoli dettati dall'ambiente d'esperienza, in virtù dei quali rilevare combinazioni di attributi frequenti (Rosch 1973; Rosch *et al.* 1976), allora sarebbe più accurata una "teoria degli esemplari" (Medin, Schaffer 1978), in grado di registrare il maggior numero di informazioni possibili di ogni singolo esemplare, senza tralasciare la progressiva variabilità dei tratti medi (Barsalou, Medin, 1986; Barsalou *et al.* 1998; Lamberts 2000; Smith, Minda 2002). Se invece i vincoli che condizionano le scelte di appartenenza si basano in ultima istanza sull'inquadramento soggettivo di caratteristiche, che possono variare in relazione al contesto (Tversky 1977; Barsalou, 1982, 1987; Lucas, Griffiths 2010), oppure derivare da teorie implicite sul mondo (Murphy, Medin 1985): possono essere privilegiati così attributi relativi, e.g., alla funzionalità e non alla similarità di un oggetto; allora in questo caso sarebbe più accurato basarsi su una prospettiva cognitiva, che consideri il primato della teoria sull'esperienza (Murphy 1993; Hampton 1993) e che contempli l'intervento di vincoli di significato causali, nella selezione e nell'immagazzinamento delle informazioni. In questa prospettiva, vale la pena considerare come alcuni studiosi si siano concentrati sulla forte interrelazione tra le informazioni derivanti dalla percezione esperienziale e le informazioni insite in concetti già posseduti, dunque in

grado di influenzare la percezione stessa del mondo (Markman, Ross 2003; Ross *et al.* 2007).

Un motivo più importante dell'inefficacia della teoria dei prototipi sul piano semantico–rappresentativo è che, qualunque sia l'origine dei vincoli di salienza implicati nei processi di categorizzazione, tali vincoli sono tradotti in liste di proprietà opportune, che funzionano come criteri di comparazione. Questo aspetto della teoria riduce effettivamente il potenziale di flessibilità espresso dalla sua impostazione anti–definitoria, soprattutto a livello semantico e descrittivo–formale (*Osherson, Smith 1981*). A dispetto della reale complessità dei rapporti di sussunzione (Griffiths *et al.* 2007) e dell'incalcolabile varietà delle combinazioni analogiche che la mente umana è in grado di operare durante i processi di organizzazione cognitiva, la semantica del prototipo si presenterebbe dunque anch'essa come riduzionistica ovvero fondata su un insieme di proprietà prestabilite.

Come noto (Fodor, Lepore 1996), tale inefficacia semantica della rappresentazione prototipale è rintracciabile soprattutto nel fatto che la selezione dei tratti tipici, mentre consente di rilevare la co–varianza degli attributi, non consente di rilevare la loro variabile correlazione o "composizionalità", che è generalmente espressa in concetti complessi, quali: "vestito da mare" oppure "lampada da giardino". Tali concetti possono rappresentare delle categorie ma poiché esprimono una combinazione di caratteristiche, che spesso appartengono a domini semantici differenti, la loro inclusione in strutture categoriche non può basarsi sulla somiglianza di tratti, né derivare dalla somma delle tipicità dei singoli costituenti. Non stupisce dunque se ciò che emerge dalla descrizione di insiemi estensionali sfumati o vaghi, all'interno di una categorizzazione a prototipi, non è una reale flessibilità bensì una certa distanza del-

le rappresentazioni dai nessi esperienziali (*Osherson, Smith 1981*; Griffiths *et al.* 2007), in cui a volte proprietà disgiunte o apparentemente contraddittorie si fondono per costituire un significato.

In conclusione, la prospettiva di Eleanor Rosch, fondata sulla possibilità che le categorie siano definite in base a tratti comuni o "somiglianze di famiglia" e peraltro sull'assunto che tali caratteristiche degli oggetti di conoscenza riflettano l'intrinseca struttura correlazionale del mondo (Rosch, Mervis 1975; Rosch *et al.* 1976), può supportare la conoscenza inferenziale di caratteristiche basiche relative ad individui "naturali" appartenenti all'esperienza comune (Goldstone, Kersten 2003); tuttavia, tale prospettiva non è in grado di spiegare l'emergenza di determinate categorie, spesso "create *ad hoc*" per soddisfare esigenze inferenziali che non sono fondate induttivamente sulla somiglianza di tratti, ma piuttosto sull'associazione o la correlazione anche "metaforica" di caratteristiche esperite (Van Dantzig *et al.* 2011; Aerts *et al.* 2015).

3. La soluzione di Barsalou

L'idea che la somiglianza non sia un criterio sufficiente ad esaurire la ricchezza del nostro potenziale cognitivo e categorizzante, ha portato ad un approfondimento dell'indagine sui processi cognitivi, che più fedelmente sembrano caratterizzare l'organizzazione della conoscenza. In questa prospettiva, la teoria della "concettualizzazione situata" di Lawrence Barsalou (Barsalou 1992, 2003, 2005, 2009, 2015) è emersa come particolarmente esplicativa e risolutoria della maggior parte delle problematiche messe in campo dai precedenti modelli. Secondo

questa teoria, i concetti non si basano su liste di tratti e non si configurano come rappresentazioni astratte, ma come schemi d'esperienza (*frames*) contenenti tutte le informazioni relative al contesto spazio–temporale di un determinato oggetto di conoscenza (Barsalou 1992). Una peculiarità di tali schemi di concettualizzazione situata è che sono rappresentazioni ricorsive ma anche temporanee: sono simulazioni provvisorie di situazioni esperite, che riproducono la varietà qualitativa di una precedente prospettiva esperienziale del percipiente (Barsalou 1992, 2009). Un'altra peculiarità di queste rappresentazioni prospettiche è che si attivano in funzione delle nuove esperienze, ma solo relativamente ad informazioni salienti: e.g. un oggetto può essere conosciuto o riconosciuto in via anticipatoria perché consistente con una o più proprietà rappresentate nello schema pertinente (Barsalou 1992, 2005, 2009).

La conoscenza dunque è fondata sull'esperienza (*grounded*) e allo stesso tempo aperta ad un'interazione continua con essa (Barsalu 1999, 2008). In quanto aperta all'interazione continua con l'esperienza, tale conoscenza è inoltre situata (*situated*) e variabile: modificando il punto di vista che si assume, cambia il modo di concettualizzare un oggetto (Barsalou 2008). In questo quadro, la categorizzazione si configura come un fenomeno realmente emergente, perché l'attivazione degli schemi consente di creare categorie basate su significati contingenti, indipendentemente dalla somiglianza percepita tra i loro membri (Barsalou 2003). In aggiunta, i nessi tra i concetti possono essere costruiti in modo diverso di volta in volta, garantendo la composizionalità e la produttività delle strutture di significato. D'altra parte, le rappresentazioni possono esprimere relazioni privilegiate, che rappresentano delle sorte di attrattori (vincoli di salienza), punti di equilibrio delle strutture situate e varianti (Barsalou 2008).

Gli schemi, dunque, consentono di rilevare sia la sistematicità (e.g. invarianza delle strutture dimensionali/formali), sia allo stesso tempo la variabilità dei tratti salienti (incluso il loro carattere correlazionale e co–variante); sono rappresentazioni coerenti, complete e allo stesso tempo flessibili, perché funzionalmente capaci di modellarsi sui contesti esperienziali.

Nel complesso, gli schemi si comportano come delle mini-teorie implicite, in grado di orientare selettivamente l'immagazzinamento di nuove informazioni, senza tuttavia determinare effetti riduzionistici sul piano conoscitivo: in questo caso i concetti e le categorie sono rappresentazioni *ad hoc* (Pezzulo 2007) che non si organizzano intorno a referenti cognitivi astratti e prestabiliti, come liste di attributi ma si costituiscono sul momento a seconda delle diverse necessità inferenziali. Nello specifico, le proprietà generali dei concetti e delle categorie che progressivamente si strutturano non sono rappresentazioni invarianti, date a priori in memoria, ma proprietà emergenti, che si determinano grazie ad una rete di combinazioni pertinenti (e variabili), prodotte e riprodotte in funzione di una memoria attiva: non una singola astrazione tipica, ma diverse astrazioni "situate" possono emergere dinamicamente a rappresentazione di una categoria. In questo, la teoria di Barsalou, pur riproponendo tutti i vantaggi di una teoria degli esemplari (Nosofksy 2011; Barsalou, 2003), esemplifica in maniera efficace la lezione wittgensteiniana ed inoltre rispetta a pieno il principio di economia cognitiva.

In conclusione, la teoria funzionalista di Barsalou, che ad oggi rimane il modello più convincente, è in grado di unificare le precedenti visioni dei processi di categorizzazione, rivalutandone la complessità. Allo stesso tempo, è in grado di conferire ai processi di categorizzazione le caratteristiche di stabilità e flessibilità classificatoria.

Tuttavia, per quanto efficace sul piano esplicativo, tale prospettiva è poco sistematica sul piano teorico: non fornisce, ad esempio, una giustificazione adeguata di come le rappresentazioni di conoscenza locali possano essere integrate all'interno di rappresentazioni di sfondo più ampie (Barsalou 2015), né di come tali rappresentazioni di sfondo, espressioni di una conoscenza cosiddetta implicita, presentino dei vincoli di natura logico–semantica costantemente validi. È difficile, infatti, pensare che la complessità rappresentativa di una struttura conoscitiva situata, fondata sempre su un edificio conoscitivo più ampio, poggi sulla memoria episodica di sistemi distribuiti (Barsalou 1999), in un'ottica puramente connessionistica; piuttosto sarebbe necessaria una spiegazione dei meccanismi in grado di caratterizzare l'origine e lo sviluppo di una memoria cosiddetta semantica. La domanda da cui partire in vista di una simile spiegazione dovrebbe riguardare il modo in cui uno schema situato sia in grado di giustificarsi come una rappresentazione di significato prospettica e intrinsecamente coerente (Barsalou 2015) ovvero: come si spiega la costituzione di uno schema (e di una concatenazione di schemi) sul piano semantico?

Secondo Barsalou (Barsalou 2008) la nozione di conoscenza fondata (*grounded*) non coincide con la nozione di conoscenza incarnata (*embodied*), nel senso che è più ampia, perché la fondazione dei processi cognitivi non è ricondotta unicamente agli stati motori bensì agli stati sensoriali in senso lato, relativi all'ambiente circostante. Le rappresentazioni di significato coinvolte non si fondano unicamente su meccanismi di codifica sensorimotoria, ma soprattutto sulla conoscenza dell'ambiente in cui tali meccanismi interagiscono e co–evolvono.

In tema di rappresentazione della conoscenza, dunque, la possibile formalizzazione di un processo cognitivo *à la*

Barsalou coinvolge non solo le nozioni di schema (Minsky 1975) e di rappresentazione percettiva (Kosslyn 1994), ma soprattutto la nozione di *grounding* (Roy 2005), in base alla quale l'acquisizione di nuovi significati non deriva dal processamento di informazioni simboliche (astratte), ma si fonda su un sostrato di significati "residenti nel mondo", che si determinano nel rapporto tra le aspettative conoscitive e il mondo stesso. Nella fattispecie, se è vero che la conoscenza "ha struttura ecologica e non del tutto astratta" (Roy 2005), qualunque tipo di funzionalismo computazionale riferito alla *grounded cognition* di Barsalou (Caruana, Borghi 2013) necessariamente si discosterà dall'anti-rappresentazionalismo spinto delle teorie enattiviste.

D'altra parte, è anche vero che qualunque riflessione sul tema della rappresentazione della conoscenza, in riferimento a questa prospettiva, non può prescindere da un esame critico delle attuali posizioni teoriche relative alla rappresentazione cognitiva fondata o *grounded*, nel suo discostarsi dal "potenziale significante" di un rappresentazionalismo cosiddetto classico.

4. Il problema del contenuto rappresentativo

L'espressione "rappresentazione concettuale", spesso utilizzata in psicologia cognitiva, esprime una complessa relazione tra pensiero, cognizione e percezione ed è per questo facilmente foriera di confusione. Si può ricondurre l'utilizzo di questa espressione alla nozione fodoriana di rappresentazione (Fodor 1975), in cui la rielaborazione simbolica di un'informazione esperienziale è promossa e integrata dalla funzione inferenziale del pensiero. La teoria rappresentazionale della mente sviluppata da Jerry Fodor

ha contribuito al rafforzamento della nozione di rappresentazione mentale (Margolis, Lawrence 2007). In particolare, ha contribuito a smorzare la netta contrapposizione imposta dal descrittivismo fregeano (Frege 1891) tra la nozione di concetto, che è dotato di un senso che sono in grado di afferrare tutti, perché obbediente all'oggettività di un'inferenza proposizionale; e la nozione di rappresentazione, che invece costituisce l'immagine mentale, infaustamente soggettiva, di un senso e di un significato. Nella prospettiva fodoriana la nozione di rappresentazione si configura come una ri–descrizione degli stati prodotti da specifici sistemi di modalità sensoriale (Fodor 1983) e come tale non viene scomposta, ma rimane un'informazione atomica, un contenuto simbolico direttamente inserito in una più vasta architettura di rappresentazioni concettuali o "categoriali" (Fodor, 1975). La sistematica produttività e composizionalità delle rappresentazioni concettuali è garantita da un rapporto inferenziale tra le singole rappresentazioni atomiche (simboli) e, nel complesso, tutti i processi cognitivi si determinano in funzione di un linguaggio del pensiero, che applica a tali simboli le stesse regole della logica proposizionale (Fodor, Pylyshyn 1988).

Nell'ambito della psicologia e delle scienze cognitive gli argomenti forniti dalla nozione fodoriana di rappresentazione hanno contribuito a rafforzare la visione del rappresentazionalismo classico ovvero di un modello della conoscenza cosiddetto *sandwich*, secondo cui i processi cognitivi costituirebbero dei meccanismi indipendenti, al centro tra percezione e azione (Hurley 2001). Gli stessi argomenti sono stati utilizzati per rafforzare la più ampia distinzione tra una conoscenza cosiddetta "amodale" ovvero solo indirettamente legata alla percezione, appunto perché mediata da rappresentazioni simboliche (Fodor, Pylyshyn

1988; Pylyshyn 1984); e una conoscenza cosiddetta "modale", che è invece essenzialmente percettiva, direttamente basata sull'attivazione e riattivazione delle modalità sensoriali (Barsalou, 1999; Barsalou *et al.* 2003) e che dunque, nel considerare la stessa esperienza senso–motoria come fonte immediata di conoscenza, concepisce la possibilità di "rappresentazioni percettive" ovvero di rappresentazioni cognitive intrinsecamente legate alle modalità con cui esperiamo (Barsalou *et al.,* 2003).

La distinzione tra rappresentazioni cognitive amodali e modali si è in certo modo sovrapposta alla distinzione tra un livello di categorizzazione basato su un atteggiamento eminentemente riflessivo e proposizionale, in grado di favorire la composizione e la comunicazione di significati e per questo definito "simbolico–concettuale"; e un livello di categorizzazione immediatamente legato alla percezione sensibile, al rapporto con lo spazio, che essendo basato su rappresentazioni di tipo analogico, non-proposizionale è stato variamente definito come "non concettuale". Il risultato è un quadro teorico poco uniforme e poco chiaro: la possibilità di afferrare/possedere un significato si ripropone a diversi livelli di astrazione cognitiva, dalla percezione sensoriale alla concettualizzazione simbolica. Non sorprendono, a questo proposito, le proposte di eliminare la nozione di "concetto" nell'ambito di un'indagine sui processi cognitivi (Gunther 2003; Machery 2009), dato il riferimento a diverse forme di rappresentazione conoscitiva, che implicano meccanismi di categorizzazione indipendenti tra loro (Machery 2009).

Per altro verso, l'indagine teorica ha inteso proporre nuove direzioni in vista di una definizione della nozione di concetto. Ne è un esempio la posizione neo–empirista di Jesse Prinz (Prinz 2002), ispirata dalle teorie di Barsalou, in base alla quale la nozione di concetto deve necessaria-

mente fondarsi sulla nozione di "rappresentazione percettiva", intesa come rappresentazione di esperienze sensoriali (Prinz 2005) e non sulla rappresentazione di simboli arbitrari. Tale posizione si traduce in un'espressa critica al rappresentazionalismo fodoriano, precisamente all'identificazione dei concetti con rappresentazioni simboliche, atomiche e amodali e al loro ruolo nella categorizzazione delle informazioni esperienziali. In questa prospettiva, l'impostazione fodoriana presenterebbe tutti gli svantaggi di un riduzionismo razionalistico, implicando la necessità di ricondurre l'esperienza ad un codice di rappresentazioni/simboli astratti, arbitrariamente fissato (Prinz 2002, 2005). Al contrario, le rappresentazioni percettive, essendo composte da diverse informazioni co–referenziali (Prinz 2005, p. 8), conterrebbero di per sé le "istruzioni" per interagire costruttivamente con altre rappresentazioni e dunque per produrre inferenze categoriali, indipendentemente da vincoli astratti. Rispetto al rappresentazionalismo classico di matrice fodoriana, il vantaggio cognitivo di una simile prospettiva (Prinz 2002; 2005) non riguarda solo il fatto che le rappresentazioni possano immediatamente coincidere con l'esperienza, ma soprattutto il fatto che la nozione di rappresentazione venga a configurarsi come un fenomeno scomponibile in un insieme aperto di nessi, che è per ciò stesso in grado di interagire produttivamente con altre componenti rappresentative.

Tuttavia, vale la pena osservare come tale critica al rappresentazionalismo classico, se per un verso contribuisce a scardinare l'ormai infecondo paradigma cognitivista basato sulla visione che i processi cognitivi e, nello specifico, i processi di categorizzazione siano costituiti da rappresentazioni simboliche computabili (Fodor, Pylyshyn 1988; Pylyshyn 1984); per altro verso non fornisce alcuna giustificazione teorica della soluzione, originariamente

proposta da Barsalou, relativa alla possibilità di comprendere significati esperienziali. Il problema infatti si ripropone: quali sono le condizioni per il possesso di una rappresentazione percettiva? In che modo si rende possibile acquisire un significato esperienziale, anche in assenza di vincoli proposizionali? E in che modo tali significati, che vanno a costituire una conoscenza cosiddetta implicita, sono in grado di interagire con livelli di conoscenza più espliciti?

Mentre il rappresentazionalismo classico, fondato sulla possibilità di manipolare simboli astratti, rimane un valido supporto alla rappresentazione formale di modelli a prototipi; manca, di contro, la teorizzazione di un adeguato rappresenazionalismo, che sia in grado di giustificare la rappresentazione formale di una conoscenza stratificata di tipo *grounded.* Pur iscrivendosi nel paradigma della conoscenza fondata e situata, prospettive come quella di Prinz, non sembrano fornire adeguato supporto teorico alla possibile rappresentazione dei processi cognitivi che in tale paradigma sono coinvolti. Ovvero manca un quadro teorico in grado di giustificare una traducibilità, sul piano logico–semantico, di tali processi cognitivi complessi, in tema di rappresentazione della conoscenza.

Probabilmente, più che verificare sul piano teorico la possibilità di una rappresentazione percettiva, che implichi tutta la ricchezza di una conoscenza implicita nonché di una conoscenza cosiddetta tacita o irriflessa, sarebbe più importante provare a teorizzare come tale conoscenza si determini. A questo proposito al fine di comprendere come si determina una rappresentazione percettiva, quale modo di conoscenza, la distinzione tra processi cognitivi modali e amodali non è forse così cruciale, quanto un'indagine sulla semantica dei contenuti rappresentativi

che a tale conoscenza conducono (Markman, Stilwell 2004).

5. La lezione del paradigma non-concettuale

A differenza che in psicologia cognitiva, focalizzata sullo studio delle strutture e dei processi di categorizzazione, l'indagine filosofica ha prevalentemente trattato la nozione di rappresentazione nel suo rapporto con la nozione di contenuto mentale. L'indagine si è articolata in maniera importante in ambito neo–fregeano, attraverso l'analisi di teorici (Evans 1982; Peacocke 1986, 1992; Dummett 1991; McDowell 1994–a), che nel tentativo di fornire una giustificazione ontologica della nozione di senso (Frege 1892) si sono soffermati sull'ipotesi che il potenziale inferenziale di un contenuto potesse essere declinato non solo in una forma concettuale astratta, ma anche in una forma percettiva e dunque non–concettuale. Secondo questa prospettiva, la percezione sensoriale non consisterebbe in un coacervo disordinato di sensazioni, ma sarebbe dotata di un contenuto rappresentazionale proprio, "afferrabile" indipendentemente dall'esercizio di concetti (Coliva 2013). L'assunto principale è che, così come i concetti (o sensi fregeani) sono portatori di un contenuto e così come tale contenuto è afferrabile in termini di possesso delle sue condizioni di correttezza (il contenuto è l'inferenza stessa: Peacocke 1992; 2008); allo stesso modo è possibile ascrivere un contenuto alle esperienze percettive in virtù di una caratterizzazione semantica dei loro requisiti di adeguatezza (Dummett 1991, Peacocke 1992). Benché indissolubilmente legati all'ambito di trattazione qui proposto, non è possibile in questa sede approfondire argomenti che riguardano lo statuto ontologico della nozione di rappresen-

tazione mentale. Interessa tuttavia, per i propositi delle presenti osservazioni, una breve riflessione sul potenziale esplicativo del paradigma non concettuale con riguardo alla determinazione di una rappresentazione di significato e, in particolare di una rappresentazione di significato esperienziale.

A questo proposito, vale la pena riflettere su alcuni argomenti dello sforzo teorico più rilevante all'interno del paradigma non-concettuale: quello di Cristopher Peacocke (Peacocke 1986, 1992). Nell'introdurre le nozioni di scenario e di proto-proposizione, Peacocke intende fornire dei criteri in grado di rendere una percezione semanticamente valutabile e dunque in grado di rendere il suo contenuto concreto un contenuto rappresentazionale, anche se non concettuale. Secondo Peacocke, con riguardo alla percezione ambientale, esistono due livelli consequenziali di rappresentazione non–concettuale: un primo livello immediato, che è ritagliato sulle nostre capacità di individuare riferimenti spaziali (contenuto di scenario) ed è nella fattispecie orientato in funzione del punto di origine dettato dalla nostra prospettiva di percipienti (scenario posizionato); un secondo livello, che ci consente una primitiva rappresentazione di relazioni semantico–funzionali tra gli elementi individuali all'interno di uno scenario (proto–proposizione), ed è fondato su capacità inferenziali simili a quelle che ci consentono la rappresentazione di una proposizione. Tale livello di interpretazione semantica non implica la discriminazione di proprietà di natura predicativa (Peacocke 1992). L'attribuzione di proprietà in forma predicativa interverrebbe ad un terzo livello di rappresentazione conoscitiva, un livello post–percettivo fondato sulla capacità di tradurre i significati esperienziali in proposizioni: un livello appunto concettuale (per intrattenere delle proposizioni sono necessari i concetti).

Come noto (Paternoster 2002), i maggiori oppositori alla tesi del contenuto non–concettuale sostengono che, se percepire significa essere in grado di inferire determinate condizioni di correttezza semantica, allora il discrimine tra percezione e pensiero concettuale difficilmente può considerarsi tracciato. Senza voler entrare nelle pieghe di un dibattito complesso e ramificato, basti considerare che il tema centrale dello scontro tra concettualismo e non-concettualismo si basa sulla possibilità di giustificare tali condizioni di correttezza ad un livello intenzionale irriflesso. Il principale argomento contro il contenuto non–concettuale è che un'inferenza può essere attivata solo da un atteggiamento proposizionale ovvero da una credenza (o acquisizione di una credenza) in grado di stabilire un referente epistemico. Ad un livello intermedio tra una simile posizione e l'intenzionalismo non-concettualista, McDowell (McDowell 1994–b) considera che il dispiegamento di abilità inferenziali sia possibile anche a livello non–proposizionale, grazie ad una sorta di adesione (*endorsement*) al contenuto percepito; ma questo sarebbe possibile solo perché, appunto, sarebbero state anticipatamente esercitate delle abilità concettuali: non sarebbe altrimenti giustificabile alcuna "adesione" ad un significato; non esistono fondamenti della conoscenza sottratti alla dimensione linguistico-concettuale (Sellars1956).

Eppure, quello che conta rilevare in merito all'intenzionalismo non–concettualista di Peacocke è quanto segue: se è vero che il referente di correttezza di un'inferenza è ciò che è costitutivo delle condizioni di possesso del relativo contenuto (il contenuto è l'inferenza stessa: Peacocke 1992; 2008 p. 54); allora nella nozione di scenario posizionato, che ci viene offerta, è la stessa prospettiva spaziale a costituire dichiaratamente (Peacocke 1992, p. 67; 2005, p.60) un primitivo criterio di correttez-

za su cui si fondano le inferenze implicate nella successiva proto-proposizione. In una prospettiva intenzionale è lo stesso scenario posizionato, che i nostri sensi ci vincolano a percepire, ad essere naturalmente e immediatamente "aderito" (senza un necessario atteggiamento proposizionale) come origine della nostra esperienza.

In aggiunta, se i costituenti di una struttura inferenziale non-concettuale non esprimono proprietà predicative (Peacocke 1992), ne consegue che la relazione tra l'elemento originario con funzione semantico–referenziale e gli altri costituenti della struttura non implicheranno vincoli di natura sequenziale e discreta, ma di natura (necessariamente) funzionale. In questa prospettiva il livello della proto-proposizione attiverebbe l'interpretazione degli elementi individuali all'interno di un possibile scenario posizionato, in funzione di una precedente adesione allo scenario stesso; cosicché ciascun elemento individuale di questo scenario esperienziale diverrebbe parte di un insieme aperto di relazioni vero-funzionali legate al referente d'origine. Tali relazioni funzionali, progressivamente articolate e innestate, verrebbero ad esprimere non solo proprietà indicali, ma anche proprietà semantiche più complesse, "fondate" sui vari stati esperienziali del percipiente.

Una simile lettura della nozione di rappresentazione non–concettuale riesce a giustificare l'applicazione di importanti assunti sul contenuto concettuale al contenuto percettivo, senza che questo sia debitore di un atteggiamento proposizionale. I più importanti di questi assunti sono: quello della composizionalità del significato (Peacocke 1992; 2008, p. 135) ovvero l'assunto che un contenuto non debba considerarsi come una rappresentazione atomica, ma di per sé come il progressivo aggregarsi di nessi costituenti; quello della non–circolarità delle condi-

zioni di possesso di un contenuto (Peacocke 1992) ovvero l'assunto che qualunque tipo di nesso inferenziale debba essere "fondato" da un referente semantico ultimo**.** Invero, tale assunto della non–circolarità implica la conseguenza interessante che tutta la struttura dei processi di conoscenza sia in ultima istanza fondata su uno stato esperienziale originario, in una prospettiva non–riduzionistica.

Così tradotta, la nozione di contenuto non–concettuale si presta a rappresentare, nonché plausibilmente a giustificare sul piano semantico–formale, i processi conoscitivi esemplificati dalla nozione di *grounded cognition* di Barsalou, quale conoscenza ricorsivamente fondata sugli stati esperienziali ambientali. Infatti, in uno scenario esperienziale l'attitudine percettiva cambia rispetto alla percezione dello scenario, perché cambia il modo in cui possono essere codificate le rappresentazioni–schemi, dunque cambia il modo in cui "le cose sono viste". La continuità semantica dei nessi esperienziali è garantita da una rete prospettica di relazioni vero-funzionali, che per un verso sono legate ai vincoli sensoriali originari, per altro verso possono variare in relazione a nuove esperienze ambientali. D'altra parte, è ipotizzabile una continuità semantica tra nessi esperienziali e inferenze post–percettive, promossa da continue articolazioni funzionali di un'unica macro–rete prospettica. In questo quadro, la stabilità e la variabilità emergente dei contenuti di conoscenza potrebbero riguardare tutti i livelli di astrazione.

A livello semantico, i vantaggi rappresentativi di una simile prospettiva deriverebbero dall'"anti–descrittivismo" di un unico imprescindibile referente originario, in grado di ancorare in maniera non riduttiva una fitta e fine rete di nessi inferenziali (dunque di rappresentazioni conoscitive) a livello implicito, nonché tacito o irriflesso. Il referente cognitivo di una rappresentazione percettiva coinciderebbe

dunque non già con un'informazione simbolica (corrispondente ad uno o più attributi oggettuali espressi in forma esplicita o predicativa), ma con un reticolato scomponibile di significati, che sono vincolati al mondo circostante e che di questa stessa rappresentazione costituiscono lo sfondo continuato di conoscenza, uno sfondo funzionalmente attivabile in relazione a diverse esigenze inferenziali e categorizzanti. Vale la pena considerare come una simile prospettiva sia in grado di supportare la rappresentazione di un funzionalismo cognitivo complesso, attraverso logiche cosiddette *default* (Reiter 1980), senza tuttavia assumere un descrittivismo dei vincoli. In questo caso, infatti, la rappresentazione di "pacchetti" di conoscenza implicita, che fungano funzionalmente da sfondo ad una conoscenza situazionale e pragmatica, sarebbe unicamente vincolata a funzioni di tipo semantico invece che ad un quadro referenziale descrittivo.

Si consideri dunque come una prospettiva come quella descritta riesca a fornire un quadro teorico di supporto ad un'adeguata rappresentazione formale di processi cognitivi *grounded*. Tale prospettiva, infatti, non solo è in grado di giustificare sul piano semantico una stratificazione della conoscenza, ma anche di supportare la rappresentazione di processi cognitivi complessi attraverso logiche non monotoniche di tipo *default*, senza incorrere nella rigidità che pure queste implicano ogni qual volta assumono un descrittivismo dei vincoli.

6. Conclusione

Le riflessioni critiche qui sviluppate hanno avuto come scopo quello di attirare l'attenzione su possibili direzioni teoriche di supporto ad una rappresentazione della cono-

scenza che abbia come riferimento processi di categorizzazione complessi, come quelli espressi dalla nozione di conoscenza fondata di Barsalou. Dalle considerazioni svolte emerge come, se è vero che un'efficace configurazione dei processi di categorizzazione non possa essere ricondotta ad una prospettiva riduzionistica (ovvero rigidamente determinata dalla descrizione dei referenti), ma debba rifarsi ad una nozione di conoscenza situata e variabile, le cui strutture di significato possano determinarsi come proprietà emergenti e possano organizzarsi in maniera funzionale e sufficientemente flessibile; allora sarà necessario ricorrere ad una prospettiva teorica in grado di giustificare tale nozione sul piano semantico-costitutivo in modo da supportare adeguatamente anche una possibile rappresentabilità logico–formale dei processi cognitivi coinvolti. A fronte di un'evidente insufficienza teorica in tale ambito, si è considerato come l'ipotesi di un certo tipo di rappresentazioni cognitive non–concettuali, fondate sull'anti–descrittivismo dei referenti semantici, possa non solo giustificare la possibilità che determinate strutture conoscitive emergano nell'esperienza (prima di una codificazione simbolica e arbitraria degli stati esperienziali), nel quadro di una conoscenza implicita o di sfondo, ma anche supportare una possibile rappresentazione formale dei legami semantici implicati. Nello specifico, questa prospettiva supporta la rappresentazione di un'inferenzialità funzionale di tipo *default*, i cui vincoli si fondano ricorsivamente su una fitta rete di legami semantici, anziché sulla codifica arbitraria di "proto–significati".

Riferimenti bibliografici

AERTS D., SOZZO S., VELOZ T., (2015) *New fundamental*

evidence of non-classical structure in the combination of natural concepts, «Philosophical transactions of the Royal Society of London» A 374-20150095.

ALXATIB S., PELLETIER J., (2011) *On the psychology of truth gaps*, in R. Nouwen, R. van Rooij, U. Sauerland, and H.-C. Schmitz (eds) *Vagueness in Communication*, Springer–Verlag, Berlin, Heidelberg, pp. 13–36.

BARSALOU L.W., (2015) *Situated conceptualization*: *Theory and application*, in Y. Coello & M. H. Fischer (eds.) *Foundations of embodied cognition*, Routledge, New York, pp. 11–37.

BARSALOU L.W., (2009) *Simulation, situated conceptualization, and prediction*, «Philosophical transactions of the Royal Society of London B: Biological Sciences» 364 (1521):1281–9.

BARSALOU L.W., (2008) *Grounded cognition*, «Annual Review of Psychology», 59:617–45.

BARSALOU L.W., *(2005) Situated conceptualization,* in Cohen, H. and Lefebvre, C. (eds.) «Handbook of Categorization in Cognitive Science». Elsevier, St. Louis, pp. 619–650.

BARSALOU L.W., (2003-a) *Situated simulation in the human conceptual system,* «*Language and Cognitive Processes*»*,* 18:513–562.

BARSALOU L.W., KYLE S.W., BARBEY A.K., WILSON C.D.*,* (2003-b) *Grounding conceptual knowledge in modality-specific systems*. «*Trends in Cognitive Sciences*», 7 (2):84–91.

BARSALOU L.W., (1999) *Perceptual symbol systems*, «*Behavioral and Brain Sciences*», 22:577–660.

BARSALOU L. W., (1992) *Frames, Concepts, and Conceptual Fields*, in Lehrer, A. & Kittay, E.F. (eds.) *Frames, Fields, and Contrasts*, Lawrence Erlbaum Associates Publishers, pp. 21–74.

BARSALOU L.W., (1987) *The instability of graded structure: implications for the nature of concepts*, in U.Neisser (ed.) *Concepts and conceptual development: Ecological and intellectual factors in categorization,* Cambridge University Press, Cambridge, pp. 101–140.

BARSALOU L.W., (1982) *Context-independent and context-dependent information in concepts*, «Memory and Cognition», 10:82–93.

BARSALOU L.W., HALE C.R., (1993) *Components of conceptual representation: From feature lists to recursive frames*, in I. Van Mechelen, J. Hampton, R. Michalski, P. Theuns (Eds.) *Categories and concepts: Theoretical views and inductive data analysis*, Academic Press, San Diego, CA, pp. 97–144.

BARSALOU L.W., MEDIN D.L., (1986) *Concepts: fixed definitions or context-dependent representations?,* «Cahiers de Psychologie Cognitive», 6:187–202.

BARSALOU L.W., HUTTENLOCHER J., LAMBERTS K., (1998) *Basing categorization on individuals and events*, «Cognitive Psychology», 36:203–272.

CARUANA F., BORGHI A.M., (2013) *Embodied Cognition: una nuova psicologia,* «Giornale Italiano di Psicologia», 1:23–48.

COLIVA A., (2013) *Dal senso ai sensi... e ritorno*, «Senso e sensibile», Prospettive tra Estetica e Filosofia del Linguaggio», vol. VII(17):63–67.

COLLINS A.M., QUILLIAN M.R., (1969) *Retrieval time from semantic Memory*, «Journal of verbal learning and verbal behavior», 8:240–248.

DUMMETT M., (1991) *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge.

EVANS G., (1982), *The Varieties of Reference* (John McDowell ed.), Oxford University Press, Oxford.

FELDMAN N.H., GRIFFITHS T.L., MORGAN J.L., (2009) *The*

influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference, «Psychological Review», 116:752–782.

FODOR J. A., *(*1983) *The modularity of mind: An essay on faculty psychology*, MIT Press, Cambridge, MA.

FODOR J. A., (1975) *The Language of Thought*, Harvard University Press, Cambridge, MA.

FODOR J.A., LEPORE E.*,* (1996) *The Red Herring and the Pet Fish*: *Why Concepts Still Can't Be Prototypes*, «Cognition», 58:253–270.

FREGE G., (1892) *Über Sinn und Bedeutung*, «Zeitschrift für Philosophie und philosophische Kritik», C:25–50.

FREGE G., (1891) *Funktion und Begriff*, «Jenaische Gesellschaft für Medizin und Naturwissenschaftjena».

GOLDSTONE R. L., KERSTEN A., (2003) *Concepts and Categories*, in A.F. Healy, R.W. Proctor (eds.) *Comprehensive handbook of psychology,* Wiley, New York, 4*:*591–621.

GRIFFITHS T. L., TENENBAUM, J.B., (2009) *Theory-based causal induction,* «Psychological Review», 116:661–716.

GRIFFITHS T.L., STEYVERS, M., TENENBAUM, J.B., (2007) *Topics in semantic representation*, «Psychological Review», 114:211–244.

GUNTHER H.Y., (ed) (2003) *Essays on nonconceptual content*, MIT Press, Cambridge, MA.

HAMPTON *J*.A., (1995) *Testing the Prototype Theory of concepts*, «Journal of Memory and Language», 34(5):686–708.

HAMPTON J., (1993) *Protoype models of concept representation*, in I. Van Mechelen, R.S. Michalski (eds) *Categories and concepts: Theoretical views and inductive data analysis,* pp. 67–95.

HAMPTON J., (1979) *Polymorphous Concepts in Semantic Memory,* «Journal of Verbal Learning and Verbal Behavior», 18 (4):441–461.

HURLEY S., (2001) *Perception and action: Alternative views*, «Synthese», 129:3–40.

KEANE M.T., EYSENCK M.W., (2005) *Cognitive Psychology: A Student's Handbook*. Routledge, Psychology Press, *Abingdon–UK*.

KOSSLYN, S., (1994) *Image and Brain: The Resolution of the Imagery Debate*, MIT Press, Cambridge, MA.

LAMBERTS K., (2000) *Information-accumulation theory of speeded categorization*, «Psychological Review», 107:227–260.

LUCAS C.G., GRIFFITHS T.L., (2010) *Learning the form of causal relationships using hierarchical Bayesian models*, «Cognitive Science», 34:113–147.

MACHERY E., (2009) *Doing without concepts*, Oxford University Press, Oxford.

MARGOLIS E., LAURENCE S., (2007) *The Ontology of Concepts -Abstract Objects or Mental Representations*?, «*Nous*», 41(4):561–593.

MARKMAN A. B., ROSS B. H., (2003) *Category use and category learning,* «Psychological Bulletin», 129:592–613.

MARKMAN A., H. STILWELL, (2004) *Concepts à la Modal: An Extended Review of Prinz's Furnishing the Mind, «*Philosophical Psychology*»* 17(3):391–401.

MINSKY M., (1975) *A framework for representing knowledge,* in P.H. Winston (ed.) *The Psycology of computer vision*, McGraw-Hill, New York, pp. 211–280.

MCDOWELL J., (1994–a) *Mind and World*, Harvard University Press, Cambridge, Mass.

MCDOWELL J., (1994–b) *The content of perceptual experi-*

ence, «*Philosophical Quarterly*», 44:190–205.

MEDIN D.L., SCHAFFER M.M., (1978) *Context theory of classification learning,* «Psychological Review», 85:207–238.

MURPHY G.L., (1993) *Theories and concept formation*, in I. Van Mechelen, R.S. Michalski (eds) *Categories and concepts: Theoretical views and inductive data analysis*, Academic Press, London, pp. 173–200.

MURPHY G.L., (2002) *The big book of concepts*, MIT Press, Cambridge, MA.

MURPHY G.L, MEDIN D.L., (1985) *The role of theories in conceptual coherence*, «Psychological Review», 92:289–316.

OSHERSON D.N., SMITH, E.E., (*1981*) *On the adequacy of prototype theory as a theory of concepts,* «Cognition», 9:35–58.

PATERNOSTER A., (2002) *Introduzione alla filosofia della mente*, Laterza, Roma–Bari.

PEACOCKE C., (2008) *Truly Understood*, Oxford University Press, Oxford.

PEACOCKE C., (*1992) A study of concepts*, *MIT* Press, Cambridge MA.

PEACOCKE C., (1986) *Thoughts: An Essay on Content*, Blackwell, Oxford.

PEZZULO G., (2007) *Rappresentazione della conoscenza*, in Bianchini F., Gliozzo A., Matteuzzi M., *Instrumentum Vocale*, Bononia University Press, Bologna.

PRINZ, J. J., (2005). *The return of concept empiricism*, in *H. Cohen*, *C. Lefebvre Handbook of Categorization in Cognitive Science*, Elsevier, Oxford, pp. 679–695.

PRINZ, J.J., (2002) *Furnishing the mind: Concepts and their perceptual basis*, MIT Press, Cambridge, MA.

PYLYSHYN Z.W., (1984) *Computation and cognition*: *Toward a foundation for cognitive science*, The MIT Press, Cambridge, Massachusetts.

REITER R., (1980) *A logic for default reasoning*, «Artificial Intelligence», 13:81–132.

ROSCH, E., (1983) *Prototype classification and logical classification: the two systems*, in E.Scholnick (ed.) *New Trends in Conceptual Representation: Challenges to Piaget Theory?*, pp. 133–159.

ROSCH, E., (1978) *Principles of Categorization*, in Rosch E., Lloyd B.B. (eds) *Cognition and Categorization*, Lawrence Erlbaum Associates, Publishers, Hillsdale, pp. 27–48.

ROSCH E.H., (1973) *Natural categories*, «Cognitive Psychology», 4:328–350.

ROSCH E., MERVIS, C.B., (1975) *Family Resemblances: Studies in the Internal Structure of Categories,* «Cognitive Psychology», 7 (4):573–605.

ROSCH E., MERVIS C.B., GRAY W.D., JOHNSON M.D., BOYES-BRAEM P., (1976) *Basic objects in natural categories*, «Cognitive Psychology», 8:382–439.

ROSS B.H., WANG R.F., KRAMER A.F., SIMONS D.J., CROWELL J.A., (2007) *Action information from classification learning*, «Psychonomic Bulletin and Review», 14:500–504.

ROY D., (2005) *Grounding words in perception and action: computational insights*, «Trends Cognitive Science», 9(8):389–396.

SELLARS W., (1956) *Empiricism and the Philosophy of Mind* (ed. Robert Brandom), Harvard University Press, Cambridge, MA.

SMITH J.D., MINDA J.P., (2002) *Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning*, «Journal of Experimental Psycholo-

gy: Learning, Memory, and Cognition», 28(4):1433–1458.

TVERSKY A., (1977) *Features of similarity*, «Psychological Review», 84:327–352.

VAN DANTZIG S., RAFFONE A., HOMMEL B., (2011) *Acquiring contextualized concepts: a connectionist approach*, «Cognitive Science», 35:1162–1189.

WITTGENSTEIN L., (2001) [1953] Philosophical Investigations, Blackwell Publishing.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/97888255130669
pp. 171-186 (marzo 2018)

Concetti e categorizzazione dei disturbi mentali

Come la psicologia cognitiva può aiutare la psichiatria

di Elisabetta Lalumera[1]

1. Introduzione

Le discipline che si occupano della ricerca e cura dei disturbi mentali si trovano da tempo, e più che mai oggi, in una fase di ripensamento metodologico e di coesistenza di vari paradigmi teorici, nel senso del termine introdotto da Thomas Kuhn (1970). Ci sono approcci psicoanalitici e psicodinamici del disturbo mentale, modelli biosociali, costruttivisti, cognitivisti-comportamentali, che convivono con il modello medico predominante. All'interno del modello medico occorre poi distinguere le spiegazioni con enfasi sulla natura genetico–biologica, neurocognitiva, oppure epidemiologico–statistica del disturbo (Ghaemi 2004; Murphy 2015). Di fronte al pericolo di una "Babele psichiatrica" (Demazeux, Singy 2015) e nel tentativo di migliorare l'affidabilità (*reliability*) o convergenza sulle diagnosi, e la comunicazione ai fini di ricerca e del rapporto con le assicurazioni sanitarie e gli *stakeholders* non medici, la American Psychiatric Association (APA) ha scelto

[1] Università degli Studi di Milano–Bicocca. E–mail: elisabetta.lalumera@unimib.it

dal 1980 di impostare il Manuale Diagnostico e Statistico dei Disturbi Mentali (DSM) – la nosologia dei disturbi mentali più usata nel mondo – con un carattere *ateorico* e *descrittivo* (APA 1980; Follette e Houts 1996; Blashfield 2012). Coerentemente con questa scelta, anche nel più recente DSM–5 (APA 2013) nessun paradigma teorico viene esplicitamente adottato (benché implicitamente si lavori all'interno di una qualche forma di modello medico), e i singoli disturbi hanno criteri diagnostici espressi da concetti criteriali descrittivi, che indicano i sintomi, la prognosi, in alcuni casi la risposta ai farmaci, ma non le cause del disturbo. Idealmente, in questo modello, la diagnosi è un procedimento di categorizzazione *check and count*, in cui il clinico controlla quanti tra i criteri diagnostici il paziente soddisfa (Westen *et al*. 2011)[2].

In psichiatria e psicologia clinica, ma anche nel dibattito filosofico su queste discipline, ci sono numerose critiche al DSM–5, sia per quanto riguarda la cosiddetta *validità* dei concetti e quindi delle categorie diagnostiche elencate (ad esempio, la dipendenza da caffeina è una patologia? La schizofrenia è veramente un unico disturbo?), sia per quanto riguarda i limiti dell'approccio ateorico e descrittivo, e quindi non esplicativo, alla caratterizzazione delle varie condizioni (Murphy 2005; Tsou 2015; Lalumera 2016). Un'altra questione rilevante per ogni nosologia, e quindi anche per il DSM, è l'*utilità clinica*, che è funzione della facilità della comunicazione tra specialisti, della facilità di uso in termini di tempo, dell'implementabilità in processi di decisione e politiche sanitarie (Reed 2010). L'ipotesi che guida il presente contributo è che uno dei

[2] Per i disturbi di personalità e per i disturbi dello spettro autistico è stata aggiunta un'appendice con criteri diagnostici dimensionali, dopo lunga discussione. Si veda ad es. WIDIGER *et al.* (2007).

modi per migliorare l'utilità clinica sia rendere il formato della nosologia quanto più simile possibile al modo in cui i clinici categorizzano i pazienti – cioè fanno diagnosi –, nell'ambito del rispetto dei vincoli di validità e coerenza interna (Mullins–Sweat *et al*. 2016). In psicologia sperimentale ci sono ricerche in corso sul modo di rappresentare concettualmente i disturbi mentali da parte dei clinici (cioè esperti) nella pratica e in contesti di laboratorio[3]. In questo breve articolo ne illustrerò una parte, suggerendo un possibile utilizzo a miglioramento dell'utilità clinica della nosologia psichiatrica corrente.

I risultati sperimentali su cui mi soffermerò rispettivamente nelle sezioni 2 e 3 sono i seguenti. Innanzitutto, i concetti dei disturbi mentali utilizzati dai clinici, sia studenti che esperti, tendono ad essere concetti-teorie (Ahn *et al*. 2000; Kim, Ahn 2002a, 2002b). Questo tipo di rappresentazione, che usiamo prevalentemente per le categorie naturali, contiene informazione strutturata, con alcune caratteristiche centrali che spiegano le altre in virtù di relazioni causali (Carey, 1985; Keil, 1989; Medin, 1989; Murphy, Medin, 1985). C'è inoltre evidenza dell'uso di rappresentazioni prototipiche dei disturbi mentali – il prototipo di una categoria è il membro ideale, cioè una rappresentazione che ne possiede le caratteristiche tipiche, cioè condivise da più membri e meno frequenti in altre categorie (Rosch 1978; Rosch, Mervis 1975; Rips *et al*. 2012). Come dirò nella terza sezione, una diagnosi per prototipi sembra essere particolarmente adatta al caso dei disturbi di personalità, in cui la conoscenza teorico–causale è meno certa a livello soggettivo e meno condivisa

[3] Le rappresentazioni in questione sono quello che i filosofi tenderebbero a chiamare "concezioni" dei disturbi mentali, ma che qui indicherò come "concetti" seguendo l'uso corrente e psicologico. Sull'uso della distinzione tra concetti e concezioni si veda LALUMERA (2014).

a livello di comunità scientifica. Complessivamente, questi studi mostrano che l'approccio del DSM non riflette la concettualizzazione dei suoi utilizzatori, e che non l'ha cambiata con la pratica (Kim, Ahn 2002a) e che altre forme di manuale diagnostico potrebbero essere più in linea con la categorizzazione dei clinici, a parità di validità nosologica (Westen *et al*. 2006).

Anticipando la conclusione (sezione 4), occorre ricordare che il modo in cui i clinici tendono a pensare alle malattie mentali non determina che cosa sia la malattia mentale, perché in generale ammettiamo che in qualche misura la scienza possa essere revisionista e prescrittiva rispetto ai nostri concetti, e anche a quelli degli scienziati occupati nella pratica di "scienza normale". Tuttavia, la psicologia dei concetti psichiatrici può dare utili indicazioni al progetto di riforma della nosologia del DSM verso una maggiore *utilità clinica*, che assieme alla validità esterna dei concetti e la coerenza delle descrizioni è uno dei criteri di adeguatezza di una classificazione.

2. Concetti–teoria

I concetti dei singoli disturbi mentali nel DSM–5 sono *criteriali* (costituiti da diversi criteri non definitori, dei quali è sufficiente soddisfare un certo numero minimo per ricevere la diagnosi), *categorici* (si applicano o non si applicano a un dato caso, senza gradazione), *descrittivi* (i criteri non contengono informazione causale) e *non gerarchici* (i criteri non hanno in generale un ordinamento dal più importante al meno importante, e la relazione fra essi non è esplicitata). Ad esempio, per ricevere una diagnosi di anoressia nervosa, uno dei Disturbi della nutrizione e

dell'alimentazione, i criteri da soddisfare sono in questa lista (APA 2013, tr. it. p. 391):

1. Restrizione dell'assunzione di calorie in relazione alle necessità, che porta a un peso corporeo significativamente basso nel contesto di età, sesso, traiettoria di sviluppo e salute fisica. Il peso corporeo significativamente basso è definito come un peso inferiore al minimo normale oppure, per bambini e adolescenti, meno di quello minimo atteso;
2. Intensa paura di aumentare di peso o di diventare grassi, oppure un comportamento persistente che interferisce con l'aumento di peso, anche se significativamente basso;
3. Alterazione del modo in cui viene vissuto dall'individuo il peso o la forma del proprio corpo, eccessiva influenza del peso o della forma del corpo sui livelli di autostima, oppure persistente mancanza di riconoscimento della gravità dell'attuale condizione di sottopeso;
4. Amenorrea, cioè assenza di almeno tre cicli mestruali consecutivi.

Per una diagnosi di Disturbo di personalità borderline un soggetto deve invece mostrare almeno cinque di questi nove comportamenti sintomatici (APA 2013, tr. it. p. 368):

1. Sforzi disperati per evitare un reale o immaginario abbandono;
2. Un pattern di relazioni interpersonali instabili e intense, caratterizzate dall'alternanza tra gli estremi di iperidealizzazione e svalutazione;
3. Alterazione dell'identità: immagine di sé o percezione di sé marcatamente e persistentemente instabili;
4. Impulsività in almeno due aree che sono potenzialmente dannose per il soggetto (es., spese sconsiderate, sesso promiscuo, abuso di sostanze, guida spericolata, abbuffate);
5. Ricorrenti comportamenti, gesti o minacce suicidari, o comportamento automutilante (*autolesionismo, tagli su braccia e gambe, bruciature di sigaretta*, ndc);

6. Instabilità affettiva dovuta a una marcata reattività dell'umore (per es., episodica intensa disforia, irritabilità o ansia, che di solito durano poche ore, e soltanto raramente più di pochi giorni);
7. Sentimenti cronici di vuoto;
8. Rabbia inappropriata, intensa, o difficoltà a controllare la rabbia (per es., frequenti accessi di ira o rabbia costante, ricorrenti scontri fisici);
9. Ideazione paranoide transitoria, associata allo stress, o gravi sintomi dissociativi.

Come si è detto sopra, il DSM prescrive che la diagnosi–categorizzazione di fronte a un nuovo caso sia una procedura *check–and–count*: il clinico controlla se la persona mostra le caratteristiche dei vari criteri e conta quanti di essi sono soddisfatti, senza considerare quali siano più importanti e quali meno, e confrontandoli con la soglia specificata dal manuale arriva a un verdetto diagnostico sì/no. Questo metodo, come si è detto, ha vantaggi e svantaggi. Da un lato dovrebbe favorire la convergenza sulla stessa diagnosi da parte di clinici di orientamento differente. D'altra parte, secondo alcuni, ha l'effetto di massimizzare la comorbidità: dato che la combinazione dei sintomi dà luogo a molti modi diversi di soffrire dello stesso disturbo, aumenta la probabilità che un singolo soggetto si qualifichi per più di una diagnosi (ad esempio depressione, disturbo da dipendenza e disturbo di personalità borderline). Benché sia plausibile che nella realtà certi disturbi tendano a co–occorrere, i critici dell'approccio criteriale del DSM sottolineano che la eccessiva comorbidità sia un effetto del metodo diagnostico (Pincus *et al*. 2004; Tsou 2015).

In che misura i concetti di psichiatri e psicologi corrispondono a questi criteri non strutturati della nosologia ufficiale? Nancy Kim, Woo–kyoung Ahn e collaboratori, in una serie di studi, hanno trovato conferme dell'ipotesi se-

condo la quale i clinici fanno piuttosto uso di *concetti–teorie*: rappresentazioni che contengono informazione causale strutturata, in cui alcuni sintomi valgono più di altri per il giudizio diagnostico, e ci sono relazioni logiche tra i criteri. In generale, nelle categorie naturali la centralità causale ed esplicativa di un tratto lo rende più saliente per la categorizzazione e più facile da ricordare (Ahn *et al*. 2000). Se un disturbo mentale è rappresentato da un concetto-teoria, un sintomo ha tanto più peso nella decisione diagnostica, quindi nella categorizzazione, quanto più è esplicativo rispetto agli altri, e in generale quanto più è connesso causalmente con altri sintomi. Kim, Ahn (2002a) hanno testato un gruppo di psicologi clinici esperti, sia un gruppo di studenti, con diversi compiti: dato un disturbo (anoressia nervosa, schizofrenia, disturbo di personalità antisociale e una specifica fobia) fornirne una caratterizzazione, disegnare una mappa concettuale specificando le relazioni tra i criteri diagnostici, e decidere sulla diagnosi di fronte a un caso ipotetico in cui siano assenti uno o più sintomi. L' "effetto causale" è stato riscontrato con valori significativi in entrambi i gruppi. Nell'anoressia, ad esempio, l'immagine corporea distorta, il rifiuto di mantenere il peso e l'amenorrea non sono considerati come una semplice lista di caratteristiche, bensì come una lista strutturata in cui il primo elemento spiega il secondo, che a sua volta spiega il terzo (Kim, Ahn 2002b, 459-61). L'effetto causale è anche associato a un effetto di tipicità (Rosch 1978): di fronte alla decisione diagnostica, i pazienti che hanno i sintomi causalmente centrali nella teoria del disturbo (ad esempio, l'umore triste nella depressione, rispetto all'aumento di peso) vengono valutati come più tipici, e aumenta la certezza soggettiva della diagnosi da parte del clinico. In un altro esperimento dello stesso studio Kim e Ahn hanno rilevato che i concetti–teorie dei soggetti par-

tecipanti tendono ad essere simili nel caso di disturbi per i quali c'è una qualche sovrapposizione tra paradigmi scientifici diversi (depressione, schizofrenia), mentre sono marcatamente differenti per i disturbi di personalità, che sono più teoricamente controversi (Kim, Ahn 2002b, 465). In generale, comunque, anche in questi casi, gli studenti come gli esperti mostrano di categorizzare mediante una teoria, anziché con la procedura *check–and–count* proposta dal manuale.

Come notano Kim e Ahn in conclusione, fare diagnosi con la propria teoria può essere la radice dell'errore diagnostico, proprio questo tipo di idiosincrasia è ciò che l'introduzione dell'approccio criteriale del DSM si proponeva di evitare; esiste infatti un'ampia letteratura sulle fallacie e i *bias* del ragionamento clinico (si veda Kim e Ahn 2002a). Tuttavia, la tendenza naturale dei clinici, come dei non esperti, a dare senso e struttura causale all'esperienza dei pazienti mediante una teoria non è solo una fallacia o scorciatoia cognitiva (Keil 2003). Di fatto su alcune micro-teorie (connessioni causali tra sintomi) c'è evidenza esterna. In senso minimale, ad esempio, nell'anoressia la distorta immagine corporea è all'origine di comportamenti di riduzione della nutrizione, e non viceversa; nella depressione maggiore potremmo dire che l'insonnia provoca stanchezza, che causa problemi di concentrazione, da cui dipende il senso di vergogna, e così via (Borsboom 2017). Se i disturbi sono almeno in parte reti di sintomi, allora un avvicinamento del criterio nosologico ufficiale alla rappresentazione concettuale dei clinici potrebbe essere fattibile: in pratica, si potrebbero esplicitare nel DSM alcune delle relazioni tra i sintomi che trovano conferma empirica, trasformando alcune dei concetti criteriali in *concetti criteriali strutturati*.

3. Prototipi

I disturbi mentali descritti dal DSM e trattati dalla psichiatria e psicologia clinica mostrano grande varietà al loro interno, da condizioni in cui è chiara la presenza di un'eziologia organica e di *biomarkers* (la malattia di Alzheimer, i disturbi ossessivo-compulsivi, il disturbo depressivo maggiore) ad altre la cui base eziopatologica è più sfuggente. A questa varietà di tipologia delle condizioni che chiamiamo "disturbi" è plausibile aspettarsi una varietà nel tipo di rappresentazioni concettuali – mentre, ricordiamo, per il DSM si tratta sempre di concetti criteriali categorici e non strutturati.

Tra le condizioni più controverse dal punto di vista della genesi e *biomarkers* ci sono i Disturbi di personalità, in particolare quelli classificati nel "tipo B": disturbo di personalità istrionico, antisociale, borderline e narcisistico, sui quali il dibattito in psichiatria è particolarmente acceso, e si è lontani da una situazione di "scienza normale" (Skodol 2012). Per la rappresentazione cognitiva e per la categorizzazione dei disturbi di personalità è stato più volte indagato il ruolo della categorizzazione per prototipi, anziché per teorie (Kendell 1975; Livesley 1991). Il prototipo, come il concetto–teoria e diversamente dalla lista non strutturata di criteri, è una rappresentazione unitaria e più cognitivamente "naturale", che può essere facilmente richiamata all'uso nel contesto diagnostico; diversamente dal concetto–teoria presuppone meno conoscenza, dato che non esplicita al suo interno relazioni causali fra i tratti[4]. Un prototipo non determina una categoria con confini

[4] Nella versione originale di Rosch le relazioni fra tratti non sono esplicitate. Per proposte più complesse sul modello della teoria dei prototipi si veda (Rips *et al.* 2012).

netti, e l'inclusione nella categoria avviene per gradi di somiglianza, con una soglia di entrata (Rips *et al*. 2012; Rosch 1978; Rosch, Mervis 1975).

In un esperimento condotto da Diana Evans e colleghi (Evans *et al.* 2002) sulla diagnosi dei disturbi di personalità ogni soggetto (psicologo clinico esperto) riceve 12 profili di pazienti, ciascuno dei quali ha caratteristiche della categoria diagnostica a tipicità alta o media, a dominanza alta o media (la dominanza qui è definita come la percentuale di caratteristiche tipiche rispetto alle caratteristiche totali), più alcuni tratti extracategoriali. Le caratteristiche – ad esempio "ha alta autostima" oppure "reagisce intensamente alle separazioni" – così come il loro *ranking* di tipicità sono derivate da studi precedenti. I risultati mostrano che sia la tipicità che la dominanza delle caratteristiche influenzano il giudizio diagnostico dei clinici, non solo il numero dei sintomi-criteri, come prescrive la modalità *check–and–count* del DSM.

In una serie di studi Drew Westen, Rebekah Bradley e colleghi hanno proposto che l'influenza dei prototipi nella diagnosi diventi un metodo diagnostico alternativo, almeno per i disturbi di personalità di tipo B. Un modo per passare dall'idiosincrasia della rappresentazione tipica di un disturbo propria del singolo terapeuta a un metodo condiviso, è quello di delineare *prototipi clinici*: intervistando un ampio campione di terapeuti, i ricercatori hanno chiesto di illustrare il loro prototipo di una data categoria diagnostica, e i risultati sono stati aggregati in forma narrativa[5]

[5] Il prototipo clinico del Disturbo di personalità antisociale è il seguente: "*Patients who match this prototype tend to be deceitful and to lie and mislead others. They take advantage of others, have minimal investment in moral values, and appear to experience no remorse for harm or injury caused to others. They tend to manipulate others' emotions to get what they want; to be unconcerned with the consequences of their actions, appearing to feel immune or*

(Westen *et al.* 2010). L'alternativa è realizzare *prototipi empirici* chiedendo ai soggetti di descrivere un paziente specifico con una determinata diagnosi, e poi analizzare statisticamente i risultati in modo da formare una rappresentazione di proprietà aggregate. Che si usi un prototipo clinico o uno empirico, il metodo diagnostico alternativo a quello criteriale richiede al terapeuta di formulare un giudizio di somiglianza tra il prototipo del disturbo e il paziente. Più in dettaglio, consiste nel valutare su una scala da 5 a 1 se il paziente esemplifica il prototipo (5), se vi corrisponde bene (4), se ne possiede caratteristiche significative (3) o meno significative (2) o nessuna (1). Westen *et al*. (2006, 2012) hanno messo a confronto la validità e l'utilità clinica delle categorie diagnostiche generate dai prototipi clinici ed empirici dei disturbi di personalità con quella delle corrispondenti categorie del DSM, ottenendo risultati a favore del metodo per prototipi in entrambi i casi. Questi lavori hanno portato a un intenso dibattito sulla riforma del DSM, che ha facilitato l'introduzione della procedura di diagnosi dimensionale come alternativa a quella criteriale standard per i disturbi di personalità, e che è tuttora in corso (Vanheule 2014).

---

invulnerable; and to show reckless disregard for the rights, property, or safety of others. They have little empathy and seem unable to understand or respond to others' needs and feelings unless they coincide with their own. Individuals who match this prototype tend to act impulsively, without regard for consequences; to be unreliable and irresponsible (e.g., failing to meet work obligations or honor financial commitments); to engage in unlawful or criminal behavior; and to abuse alcohol. They tend to be angry or hostile; to get into power struggles; and to gain pleasure or satisfaction by being sadistic or aggressive toward others. They tend to blame others for their own failures or shortcomings and believe that their problems are caused entirely by external factors. They have little insight into their own motives, behavior, etc. They may repeatedly convince others of their commitment to change but then revert to previous maladaptive behavior, often convincing others that 'this time is really different'." (WESTEN *et al.* 2006, 847).

4. Conclusioni

In questo breve contributo di rassegna ho illustrato come ricerche recenti abbiano evidenziato l'uso di concetti-teoria e di prototipi per i disturbi mentali da parte dei terapeuti, sia in contesti di diagnosi che condizioni sperimentali, laddove il DSM presenta un'uniforme caratterizzazione criteriale non strutturata. La psicologia cognitiva può aiutare la psichiatria e in generale le discipline che si occupano della terapia dei disturbi mentali indagando i concetti e le procedure di categorizzazione dei clinici nella pratica diagnostica. I concetti o concezioni degli esperti non sono di per sé validi – la validità va accertata su base empirica –; una volta validati, tuttavia, possono essere considerati come guida o correttivo per la compilazione di un manuale diagnostico che raggiunga un risultato di utilità clinica migliore del DSM attualmente in uso. Almeno per quanto riguarda lo strumento per la diagnosi, la Babele psichiatrica potrebbe avere un rimedio nell'accordo sulle rappresentazioni cognitive.

Riferimenti bibliografici

AMERICAN PSYCHIATRIC ASSOCIATION (APA), (1980) *Diagnostic and statistical manual of mental disorders (3rd ed.)*, Washington, DC, American Psychiatric Publications.

AMERICAN PSYCHIATRIC ASSOCIATION (APA), (2013) *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. Washington, DC: American Psychiatric Publications, tr. it. *DSM-5: Manuale diagnostico e statistico dei disturbi mentali*, Milano: R. Cortina, 2014.

AHN, W. ET AAAAL. (2000) *Causal status as a determinant*

of feature centrality, «Cognitive Psychology» 41, 361–416

AHN, W., FLANAGAN, E., MARSH, J., SANISLOW, C., (2006) *Beliefs about essences and the reality of mental disorders*. «Psychological Science», 17, 759–766.

BLASHFIELD, R., (2012) *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches*. Dordrecht, Springer Science & Business Media.

BORSBOOM, D., (2017) *A network theory of mental disorders*, «World psychiatry», 16(1): 5–13.

CANTOR, N., SMITH, E. E., FRENCH, R. D., MEZZICH, J. (1980). *Psychiatric diagnosis as prototype categorization*, «Journal of Abnormal Psychology», 89(2), 181–193.

CAREY, S., (1985) *Conceptual change in childhood,* Cambridge, MA, Plenum.

FOLLETTE, W. C., HOUTS, A. C., (1996) *Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM*, «Journal of Consulting and Clinical Psychology», 64(6), 1120.

HAMPTON, J. A. (2006) Concepts as prototypes, «Psychology of Learning and Motivation», 46, 79–113.

KEIL, F. C., (1989) *Concepts, kinds, and cognitive development.* Cambridge, MA, MIT Press.

KEIL, F. C. (2003) *Folkscience: Coarse interpretations of a complex reality*, «Trends in cognitive sciences», *7*(8), 368–373.

KENDELL, R. E., (1975) *The role of diagnosis in psychiatry*. Oxford, Blackwell Scientific Publications.

KIM, N. S., AHN, W. K. (2002a) *The influence of naive causal theories on lay concepts of mental illness*, «The American journal of psychology» 115(1), 33.

KIM, N. S., AHN, W., (2002b) *Clinical psychogists' theory–based representations of mental disorders predict their diagnostic reasoning and memory*, «Journal of Experimental Psychology: General», 131, 451–476.

KUHN, T. (1962/1970) *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, tr. it. *La struttura delle rivoluzioni scientifiche*, Torino, Einaudi.

LALUMERA, E., (2014) *On the explanatory value of the concept–conception distinction*, «Rivista Italiana di Filosofia del Linguaggio», 8(3), 73–81.

LALUMERA, E., (2016) *Saving the DSM-5? Descriptive conceptions and theoretical concepts of mental disorders*, «Medicina & Storia», 109-128.

LIVESLEY, W. J. (1991) *Classifying personality disorders: Ideal types, prototypes, or dimensions*? «Journal of Personality Disorders»*, 5*, 52-59.

MEDIN, D. L. (1989) *Concepts and conceptual structure*. «American Psychologist»*,* 12*,* 1469–1481.

MURPHY, D. (2005) *Psychiatry in the scientific image*, Harvard: MIT Press

MURPHY, D. (2015) *Philosophy of psychiatry*, in *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/archives/spr2017/entries/psychiatry/

MURPHY, G. L., MEDIN, D. L. (1985) *The role of theories in conceptual coherence*, «Psychological Review»*,* 92*,* 289–316.

PINCUS, H. A., TEW JR, J. D., FIRST, M. B. (2004) *Psychiatric comorbidity: is more less*? «World Psychiatry», 3(1): 18.

REED G.M. (2010) *Toward ICD-11: improving the clinical utility of WHO's International Classification of Mental*

Disorder, «Professional Psychology Research and Practice», 41: 457–64.

RIPS, L. J., SMITH, E. E., MEDIN, D. L. (2012) *Concepts and categories: Memory, meaning, and metaphysics*, in *The Oxford handbook of thinking and reasoning*, Oxford, OUP, 177–209.

ROSCH, E. (1978) Principles of categorization, in E. Rosch, B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27– 48), Hillsdale, NJ, Erlbaum.

ROSCH, E., MERVIS, C. B. (1975) *Family resemblances: Studies in the internal structure of categories*, «Cognitive Psychology», 7, 573–605.

SKODOL, A. E. (2012) *Diagnosis and DSM–5: Work in progress*, in *The Oxford handbook of personality disorders*, Oxford: OUP, 35–57.

SLOMAN, S. (2005) *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.

SPITZER, R. L., FIRST, M. B., SHEDLER, J., WESTEN, D., SKODOL, A. E., (2008) *Clinical utility of five dimensional systems for personality diagnosis: a "consumer preference" study*, «The Journal of nervous and mental disease» 196(5): 356–374.

TSOU, J. Y. (2015) *DSM-5 and psychiatry's second revolution: Descriptive vs. theoretical approaches to psychiatric classification*, in Steeves Demazeux, Patrick Singy (eds.), *The DSM–5 in Perspective*, Amsterdam, Springer Netherlands, 43–62.

VANHEULE, S. (2014) *Diagnosis and the DSM: A critical review*. Amsterdam, Springer.

WESTEN, D., SHEDLER, J., BRADLEY, R. (2006) *A prototype approach to personality disorder diagnosis*, «American Journal of psychiatry» 163(5): 846-856.

WESTEN, D., DEFIFE, J. A., BRADLEY, B., HILSENROTH, M.

J. (2010) *Prototype personality diagnosis in clinical practice: A viable alternative for DSM–5 and ICD–11,* «Professional Psychology: Research and Practice», 41(6): 482.

WESTEN, D., SHEDLER, J., BRADLEY, B., DEFIFE, J. A. (2012) *An empirically derived taxonomy for personality diagnosis: Bridging science and practice in conceptualizing personality*, «American Journal of Psychiatry» 169(3): 273-284.

WIDIGER, T. A., SIROVATKA, P. J., REGIER, D. A., SIMONSEN, E. (Eds.). (2007) *Dimensional models of personality disorders: Refining the research agenda for DSM-V*, Washington, American Psychiatric Publications.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/978882551306610
pp. 187-209 (marzo 2018)

Categorizzare non vuol dire solo classificare

Alcune riflessioni sui limiti dell'indagine sperimentale sulla categorizzazione

di Stefania Moretti, Alberto Greco[1]

1. Introduzione

Per fare chiarezza tra i dati dell'esperienza e ridurre la complessità ambientale ci serviamo di processi mentali in grado di registrare le informazioni che riceviamo dall'esterno, all'interno di costrutti organizzati. Questi costrutti, le categorie, assolvono la preziosa funzione di guidare il nostro comportamento, permettendoci di classificare nuove informazioni sulla base di quelle già acquisite e di integrare le nuove conoscenze con quanto abbiamo incontrato ed elaborato in passato. La categorizzazione, quindi, è quel fenomeno cognitivo complesso in cui entrano in gioco sia processi di acquisizione che di uso delle conoscenze.

Nell'ambito della psicologia cognitiva, le ricerche sperimentali condotte sulla categorizzazione si sono concentrate principalmente sul secondo tipo di processi, tralasciando tutti quei meccanismi che intervengono nella formazione delle categorie.

[1] Cognilab, Laboratorio di Psicologia e Scienze Cognitive, Università degli Studi di Genova. E–mail: stefania.moretti@edu.unige.it; greco@unige.it.

Il presente contributo ha lo scopo di sottolineare l'importanza di un'indagine che tenga conto non solo di come classifichiamo nuovi esempi sulla base di categorie apprese ma anche di come le categorie vengono formate a partire dai loro casi particolari.

La struttura della trattazione è la seguente: nella prima parte verranno introdotti i principali modelli formulati in psicologia cognitiva sulla categorizzazione e ne saranno evidenziati i principali limiti teorici; nella seconda parte si analizzerà come la limitata portata esplicativa di questi modelli derivi anche dai limiti metodologici dei paradigmi sperimentali impiegati, con particolare riferimento ai compiti di classificazione; successivamente si opererà un confronto con i metodi utilizzati in alcuni ambiti dell'Intelligenza Artificiale interessati all'apprendimento e alla rappresentazione delle conoscenze; nella parte finale verrà delineata una possibile soluzione alla questione, attraverso la presentazione di un paradigma sperimentale innovativo in grado di rilevare i processi cognitivi messi in atto durante la formazione delle categorie; in conclusione, si discuteranno le implicazioni teoriche e i possibili usi del nuovo metodo, attraverso una riflessione sulla varietà e complessità dei processi che intervengono nella categorizzazione.

2. I principali modelli della categorizzazione in psicologia cognitiva

Da più di 50 anni la psicologia cognitiva tenta di comprendere in che modo le persone organizzano le conoscenze all'interno delle categorie, attraverso indagini condotte in laboratorio.

La maggior parte di queste ricerche sperimentali studia i processi di categorizzazione attraverso compiti di classificazione, in cui ai partecipanti viene richiesto di determinare l'appartenenza o meno di nuovi esempi a una categoria data.

I modelli teorici sviluppati sulla base della performance dei soggetti partecipanti in questo tipo di compiti possono essere suddivisi in tre classi principali (Markman, Ross 2003): i modelli basati sulle regole, quelli basati sugli esemplari e quelli basati sui prototipi.

Il primo tipo di modelli fa riferimento a quella che è stata definita la "teoria classica" della categorizzazione, secondo cui una categoria viene rappresentata mentalmente dall'insieme stabile delle sue proprietà definienti, e cioè dalle caratteristiche singolarmente necessarie e congiuntamente sufficienti che la identificano (Bruner, Goodnow, Austin 1956; Ashby, Maddox 1998, 2005). Secondo questa teoria la categorizzazione consisterebbe nel verificare che i casi incontrati possiedano tutte le caratteristiche per appartenere a una certa classe. In altri termini, quando nel compito di classificazione viene presentato un nuovo stimolo da classificare, il processo messo in atto dal soggetto partecipante è un'operazione di confronto tra le proprietà dell'esemplare e l'insieme delle proprietà che definiscono la categoria appresa. Questa lista di proprietà verrebbe astratta e codificata sotto forma di regola durante la fase di apprendimento, quando vengono osservati i diversi membri della categoria. Alla base di questo tipo di modelli, dunque, si trova l'ipotesi che le scelte categoriali si basino sull'astrazione di una regola che definisce le proprietà di una categoria.

Stando ai modelli basati sugli esemplari (Brooks 1978; Medin, Schaffer 1978; Hintzman 1986; Nosofsky 1986; Estes 1986, 1994; Kruschke 1992; Lamberts 2000), inve-

ce, le scelte categoriali si baserebbero sul confronto tra i nuovi esemplari e la rappresentazione mnestica di ogni membro della categoria. Alla base della categorizzazione, in questo tipo di modelli, non viene posto un processo di astrazione ma un processo di calcolo della similarità tra il nuovo caso e tutti i casi particolari osservati durante l'apprendimento della categoria. Nella classificazione, inoltre, questi casi particolari immagazzinati in memoria verrebbero ripescati nella globalità delle loro caratteristiche. La decisione circa l'appartenenza di un esemplare a una categoria, quindi, non viene affidata a una definizione astratta ma a un giudizio di similarità globale: se l'esemplare da classificare è simile ai membri precedentemente osservati, allora verrà considerato anch'esso un membro della categoria.

Dunque, l'astrazione posta alla base del primo tipo di modelli implica l'omissione di tutto ciò che non è stabilito dalla regola categoriale, mentre i modelli basati sugli esemplari sono in grado di spiegare come quell'informazione omessa venga spesso utilizzata nei compiti di classificazione per determinare l'appartenenza di un esemplare a una categoria.

Un approccio alternativo è quello dei modelli basati sui prototipi (Posner, Keele, 1968, 1970; Reed 1972; Rosch 1973, 1975; Homa, Sterling, Trepel 1981; Smith, Minda 1998, 2001), secondo cui un nuovo esemplare viene confrontato soltanto con alcuni membri della categoria, quelli più tipici. Questi membri, i prototipi, che possiedono le caratteristiche più rappresentative della categoria, possono essere degli esemplari specifici selezionati tra i membri oppure degli esemplari ideali astratti durante l'apprendimento. Di conseguenza, durante la classificazione, il confronto avviene tra il nuovo esemplare e la rappresentazione mnestica del prototipo. Anche in questo tipo

di modelli i giudizi categoriali si basano su un processo di similarità: il grado di appartenenza di un esemplare alla categoria viene determinato in base al grado di similarità con il prototipo.

Questi tre tipi di modelli, presi singolarmente, tuttavia, non sono in grado di rendere conto completamente dei dati ottenuti sperimentalmente sulle performance categoriali (Murphy 2002; Markman, Ross 2003), ed esiste un notevole dibattito (nonché infruttuoso: vedi Gagliardi 2009) su quale sia il modello della categorizzazione più corretto.

Al di là dei limiti specifici dei singoli modelli, si possono individuare alcune criticità importanti che riguardano tutti e tre i tipi di approccio e che ne limitano la portata esplicativa.

Un primo aspetto critico riguarda l'assunto centrale che la classificazione sia la funzione primaria della categorizzazione: l'apprendimento attraverso la classificazione viene considerato come il veicolo principale della formazione delle categorie. Tuttavia, si sa che le categorie possono essere apprese nelle circostanze più disparate ed è stato dimostrato come la loro identità vari a seconda del tipo di compito categoriale richiesto (Whittlesea, Brooks, Westcott 1994; Markman, Yamauchi, Makin 1997; Yamauchi, Markman 1998; Markman, Ross 2003). Queste dimostrazioni hanno evidenziato che l'apprendimento di categorie non è mediato da un unico meccanismo categoriale ma da processi multipli e qualitativamente distinti, e che l'abilità di determinare l'appartenenza di un nuovo esemplare a una categoria è senz'altro un'importante funzione della categorizzazione ma non l'unica (Ross, Murphy 1999; Ashby, Maddox 2005).

Un'altra questione critica riguarda, nello specifico, il tipo di rappresentazione che i diversi modelli pongono alla base dei giudizi categoriali, e, cioè, il termine di paragone

che viene usato per decidere se un nuovo caso appartiene o meno a una categoria. Queste rappresentazioni interne, siano esse regole, esemplari o prototipi, vengono assunte teoricamente più che essere dimostrate empiricamente. Ogni modello propone una diversa tipologia di rappresentazione e le proprietà di queste rappresentazioni si basano tipicamente su un postulato teorico (Tunney, Fernie 2012).

In sintesi, quindi, i vari modelli della categorizzazione fanno riferimento a un unico aspetto del processo categoriale, quello classificatorio, che è quello maggiormente indagato nelle tradizionali ricerche sperimentali di psicologia cognitiva. Al contrario, i processi che intervengono nella formazione delle categorie vengono regolarmente inclusi come assunti teorici nei vari modelli ma esclusi o tralasciati dalle indagini sperimentali.

A questo punto, è rilevante considerare anche alcune questioni critiche che riguardano nello specifico gli aspetti metodologici dei compiti di classificazione.

3. Limiti metodologici degli attuali paradigmi sperimentali

La maggior parte delle ricerche sulla categorizzazione ha adottato una procedura sperimentale che prevede due fasi: una fase di training, in cui i soggetti sperimentali imparano a classificare gli esemplari di una o più categorie, e una fase di test, a partire dalla quale gli sperimentatori valutano il tipo di rappresentazioni che i soggetti si sono formasti. Durante il training, in genere, gli esemplari vengono mostrati sequenzialmente e viene chiesto di classificarli in una o più categorie. In questa fase di apprendimento, che è generalmente supervisionato, ai soggetti partecipanti viene dato un feedback dopo ogni prova in modo tale da far ap-

prendere le categorie per prove ed errori. Successivamente, quando le categorie sono state apprese con un certo grado di accuratezza, ai partecipanti vengono mostrati nuovi esemplari per i quali viene richiesto di determinare la categoria di appartenenza. In questa fase, detta di *transfer*, cioè di trasferimento delle conoscenze apprese durante la fase di training, generalmente, non viene fornito nessun feedback.

Sebbene questa procedura sia diventata ormai uno standard nella maggior parte delle ricerche sperimentali sulla categorizzazione, è possibile individuare almeno tre aspetti critici circa il modo di valutare la performance categoriale in questo tipo di test (Moretti, Greco 2016).

Un primo limite riguarda la scelta degli stimoli di trasferimento, cioè degli esempi mostrati nella fase di test. Questi item vengono scelti e costruiti appositamente dallo sperimentatore per acquisire informazioni circa il tipo di rappresentazione formatasi durante la fase di apprendimento. Di conseguenza, il materiale adoperato non è neutro ma predeterminato dallo sperimentatore e questo costituisce un problema perché il modo in cui gli stimoli di trasferimento vengono costruiti può, da un lato, dipendere dall'ipotesi che lo sperimentatore vuole testare, e dall'altro, può influenzare i giudizi categoriali dei partecipanti. La scelta di particolari stimoli, con determinati attributi, può, infatti, favorire una strategia categoriale a discapito di un'altra, o incoraggiare un tipo di strategia diversa da quella che si sarebbe usata per apprendere le categorie (Donkin, Newell, Kalish, Dunn, Nosofsky 2015).

In aggiunta a questo tipo di problematica, è possibile riscontrare alcune criticità anche nella tecnica implementata per analizzare le scelte categoriali dei partecipanti. Nei compiti di classificazione, le rappresentazioni formatesi durante l'apprendimento sono inferite dagli sperimentatori

attraverso un'operazione di confronto tra gli item di trasferimento e gli item mostrati nel training. In particolare, questo confronto viene calcolato in termini di presenza (o assenza) delle caratteristiche osservate nel training negli item che i partecipanti hanno indicato come appartenenti (o meno) alle categorie apprese. Di conseguenza, inferire le rappresentazioni categoriali a partire dalle scelte categoriali operate nel test di trasferimento risulta essere una procedura di analisi di tipo "indiretto", e criticabile, perché mediata dagli stimoli di trasferimento.

Inoltre, come è stato fatto notare dalla letteratura critica sull'argomento (per es. Donkin *et al*. 2015), le conclusioni teoriche circa il tipo di strategia categoriale adottata sembrano dipendere fortemente dalle tecniche di analisi utilizzate per individuare tali strategie. Quindi, la forza delle conclusioni a cui arrivano i vari modelli della categorizzazione dovrebbe essere sempre considerata in relazione al tipo di analisi implementata.

Un ultimo limite individuabile riguarda la valutazione degli aspetti qualitativi della performance categoriale. Raramente nei compiti di classificazione viene indagata la qualità del criterio di classificazione utilizzato dai partecipanti come, ad esempio, il livello di accuratezza dello stesso o il grado di consapevolezza dei partecipanti nel compiere le scelte categoriali. Questo tipo di informazioni risulta invece essere importante per l'indagine sui processi di categorizzazione, proprio perché è possibile che si operino delle buone scelte categoriali ma seguendo un criterio approssimativo, o non essendone pienamente consapevoli.

In sintesi, quindi, nei compiti di classificazione, dal momento che la performance viene valutata esclusivamente sulla base delle scelte operate nel test di trasferimento, si perdono tutte quelle informazioni che riguardano la qualità della rappresentazione categoriale raggiunta. Questo

tipo di informazioni si potrebbe ottenere, ad esempio, attraverso la richiesta della verbalizzazione esplicita del criterio adoperato nei giudizi di classificazione o attraverso la valutazione del livello di fiducia dei partecipanti nel compiere le scelte.

Nella letteratura sulla categorizzazione sono state avanzate diverse soluzioni alternative ai compiti di classificazione, senza però risolverne in modo soddisfacente le criticità. Ad esempio, sono stati proposti compiti di inferenza delle caratteristiche, in cui non viene richiesto di indicare la categoria di appartenenza di un nuovo stimolo ma di individuare la caratteristica mancante in uno stimolo di cui è resa nota la categoria (Yamauchi, Markman 1998; Markman, Ross 2003; Johansen, Kruschke 2005; Nilsson, Olsson 2005; Hoffman, Rehder 2010).

Al di là del dibattito sul come l'inferenza di caratteristiche sia un processo formalmente identico a quello della classificazione (Markman, Ross 2003), resta anche per questo compito il problema degli stimoli preselezionati dallo sperimentatore. Analogamente, risulta poco efficace il tentativo di migliorare i compiti di classificazione monitorando, ad esempio, il movimento degli occhi dei partecipanti durante la fase di trasferimento (Richardson, Spivey 2000; Richardson, Kirkham 2004; Scholz, von Helversen, Rieskamp 2015). Anche se questo tipo di tecnica permette di ricavare informazioni aggiuntive, i processi indagati sono esclusivamente quelli implementati nella classificazione di nuovi esemplari. I processi che, invece, intervengono nella formazione delle categorie continuano ad essere analizzati solo "indirettamente", e cioè a partire dalle strategie di classificazione con gli stimoli di trasferimento.

L'impossibilità, quindi, di discriminare sperimentalmente tra i due tipi di processi, quelli di acquisizione e formazione di una categoria e quelli di classificazione di

nuovi esempi sulla base delle categorie apprese, risulta essere un problema dell'attuale indagine sulla categorizzazione in psicologia cognitiva.

Nell'ambito dell'Intelligenza Artificiale, invece, un altro campo storicamente interessato alla rappresentazione delle conoscenze (Russel, Norvig 2002), la distinzione tra i due tipi di processi è posta alla base dello sviluppo dei sistemi di classificazione artificiale.

4. La categorizzazione nell'ambito dell'Intelligenza Artificiale

L'Intelligenza Artificiale è la disciplina che si occupa dello sviluppo di sistemi computazionali in grado di esibire comportamenti intelligenti, come quelli degli esseri umani. Una branca di questa disciplina, il settore dell'apprendimento automatico (Duda, Hart, Stork 2001; Langley 1986; Michie, Spiegelhalter, Taylor 1994; Witten, Frank 2005), si dedica allo studio di sistemi classificatori capaci di apprendere dall'esperienza. Tra questi esistono dei sistemi che usano l'*instance–based learning* (Cover, Hart 1967; Aha, Kibler, Albert 1991), una tecnica attraverso cui le categorie sono apprese a partire dai loro casi particolari. Nello specifico, fornito un training set di esemplari, cioè un insieme degli esemplari di una o più categorie, questo tipo di sistemi è in grado di estrapolare delle rappresentazioni che verranno usate successivamente per classificare nuove istanze. In questo ambito, l'abilità di acquisizione e quella di uso delle conoscenze corrispondono a due algoritmi distinti: l'algoritmo di apprendimento permette di estrarre da un *data–set* un insieme di rappresentazioni; l'algoritmo di classificazione, invece, permette di assegnare un'etichetta categoriale a ogni nuova

istanza attraverso un confronto con le rappresentazioni estratte. Le due fasi, di apprendimento e di classificazione, quindi, sono implementabili e analizzabili separatamente (Gagliardi 2014).

Nella prefazione di un libro sul *data mining* (Witten, Frank 2005), che è quell'insieme di tecniche che permettono a un sistema artificiale di estrarre conoscenza a partire da un insieme di dati, gli autori spiegano come studiare questo tipo di procedura sia importante, nell'ambito del *machine–learning*, non solo per ottenere buone classificazioni ma anche per ricavare informazioni sul tipo conoscenza estratta:

> We interpret machine learning as the acquisition of structural descriptions from examples. The kind of descriptions found can be used for prediction, explanation, and understanding. Some data mining applications focus on prediction: forecasting what will happen in new situations from data that describe what happened in the past, often by guessing the classification of new examples. But we are equally—perhaps more—interested in applications in which the result of "learning" is an actual description of a structure that can be used to classify examples. This structural description supports explanation, understanding, and prediction. (Witten, Frank 2005, p. XXIV);

e più avanti nel libro:

> As well as performance, it is helpful to supply an explicit representation of the knowledge that is acquired. In essence, this reflects both definitions of learning considered previously: the acquisition of knowledge and the ability to use it. [...] Experience shows that in many applications of machine learning to data mining, the explicit knowledge structures that are acquired, the structural descriptions, are at least as important, and often very much more important, than the ability to perform well on new examples. People frequently use data mining to gain knowledge, not just predictions (*ibidem*, p. 9).

Sebbene lo studio di come i sistemi artificiali apprendono abbia, nell'ambito del *machine–learning*, finalità principalmente applicative, in queste pagine gli autori riflettono sul valore dell'apprendimento indipendentemente dalla sua funzione pratica classificatoria. Per gli autori apprendere significa sia acquisire conoscenza che usarla; tuttavia, estrarre conoscenza da un insieme di osservazioni viene considerata un'abilità fondamentale e fondante rispetto alla capacità di classificare nuovi esemplari. Per riuscire a predire la categoria di un caso particolare, infatti, è necessario prima di tutto che si sia acquisito un certo tipo di conoscenza su quella categoria, ma non solo: l'estrazione di conoscenza è un processo che è alla base, oltre che della predizione, anche della comprensione e della spiegazione. Conoscere, quindi, riflettono gli autori, non serve esclusivamente a predire cosa accadrà nel futuro sulla base di quanto esperito nel passato ma anche a comprendere e a esplicitare quello che si sta esperendo. Inoltre, aggiungono, le persone spesso usano il *data mining* non per fare previsioni ma anche solamente per acquisire conoscenza. Nell'ambito dell'Intelligenza Artificiale, quindi, si ha ben presente sia la distinzione tra le due abilità di acquisizione e uso delle conoscenze che sono coinvolte nell'apprendimento, sia l'importanza di dare rilievo alla prima rispetto a quanto si è fatto, da sempre, con la seconda.

Lo stato dell'indagine sui processi categoriali, in un settore interessato alla simulazione dei processi cognitivi umani, quindi, funge da importante termine di paragone per la situazione dell'attuale ricerca sulla categorizzazione in psicologia cognitiva. Alla luce di questo confronto, appare sempre più chiara la necessità di ideare paradigmi sperimentali alternativi che permettano di analizzare e stu-

diare la categorizzazione indipendentemente dalla sua funzione classificatoria, o almeno non limitandola ad essa.

Nel prossimo paragrafo verrà descritto un compito di categorizzazione innovativo che è stato progettato con lo scopo di superare, almeno in parte, i limiti dei tradizionali compiti di classificazione.

5. L'Active Feature Composition task

L'*Active Feature Composition* (AFC) *task* (Greco, Moretti 2017) è un test che permette di ricavare informazioni circa il tipo di processi cognitivi messi in atto durante la formazione di categorie. Si tratta di un compito di produzione in cui i partecipanti non devono classificare nuovi esemplari ma sono coinvolti attivamente nella produzione di item che ritengono appartenenti alle categorie apprese.

La procedura è la seguente [Fig. 1]: alla fine del training, dopo l'osservazione di un insieme di casi particolari di una o più categorie, ai partecipanti viene richiesto di selezionare delle caratteristiche da un set contenente tutte le dimensioni possibili degli attributi delle categorie apprese, per combinarle insieme e ottenere così l'esempio completo. Una volta completata la combinazione, i partecipanti devono assegnare a ciascun esemplare assemblato un'etichetta categoriale, indicandone così la categoria di appartenenza.

La Figura 1 descrive schematicamente le differenze tra il classico compito di classificazione e l'Active Feature Composition task: ciò che viene richiesto di fare in un tradizionale compito di classificazione è crearsi una nuova rappresentazione per ciascuno degli item di trasferimento e poi *confrontare* queste rappresentazioni con la rappresentazione della categoria creata durante il training, per

verificare la loro appartenenza alla categoria X. Questa procedura implica quindi una doppia rappresentazione. Nel compito AFC, invece, si richiede semplicemente di *usare* la rappresentazione della categoria X, creata a partire dagli esemplari del training, per produrre esempi che si ritengono appartenenti a tale categoria.

Un'ulteriore differenza tra i compiti di classificazione e l'AFC task risiede nel fatto che per classificare devono essere necessariamente usate delle etichette linguistiche e tali etichette sono assegnate agli esempi nel loro complesso. Nel compito di produzione AFC, invece, le etichette linguistiche vengono assegnate soltanto ad esemplare assemblato, di conseguenza questo procedimento di produzione non può influenzare direttamente l'assegnazione categoriale.

Come è chiaro, dunque, i processi cognitivi coinvolti nei due diversi compiti, di produzione e di classificazione, sono diversi: la classificazione richiede un processo di confronto tra la rappresentazione della categoria acquisita e quella dei singoli item di trasferimento; al contrario, la produzione richiede di esplicitare la rappresentazione interna formatasi durante l'apprendimento, attraverso la costruzione di esemplari che si ritengono appartenenti alla categoria appresa. Di conseguenza, le scelte categoriali dipendono esclusivamente dalla rappresentazione della categoria ed è proprio a partire dagli item prodotti dai partecipanti che è possibile analizzarla in modo diretto.

Un primo tipo di analisi consiste nel confronto tra gli item prodotti e quelli mostrati durante il training. Questa analisi può avvenire sugli esempi completi o, separatamente, sulle dimensioni delle singole caratteristiche scelte. Confrontando gli esempi creati, analizzati come combinazioni di caratteristiche, è possibile calcolare il grado di similarità tra questi e gli esemplari osservati; è possibile,

cioè, calcolare se e quanti esemplari sono stati creati esattamente uguali a quelli del training (ripescando in memoria gli esemplari specifici incontrati); oppure se si sono riprodotti solo determinati esemplari osservati (ad esempio, quelli più tipici) o, diversamente, se ne sono stati creati di nuovi mai osservati (ad esempio, seguendo una regola astratta).

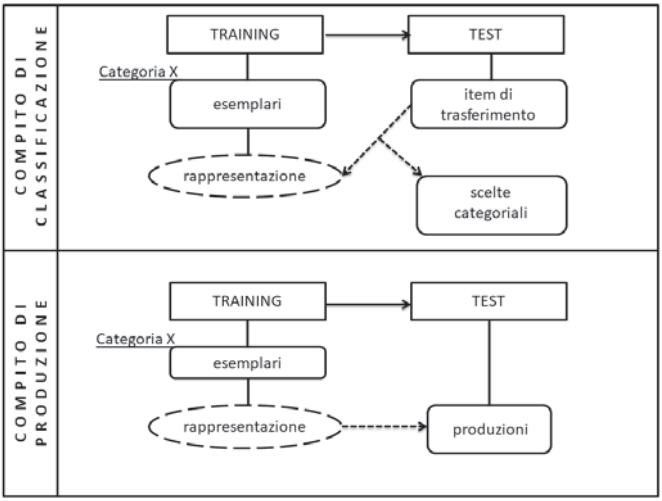

Figura 1. Differenze tra compito di classificazione (standard) e compito di produzione (*Active Feature Composition Task*)

L'analisi separata delle singole caratteristiche scelte, invece, permette di individuare quali dimensioni sono state percepite come criteriali e quali come irrilevanti, attraverso il calcolo della frequenza con cui sono state scelte per produrre i vari item.

Analizzando le caratteristiche criteriali, ad esempio, è possibile valutare il grado di accuratezza e di completezza della rappresentazione, individuando se sono state inserite in ogni item prodotto, oppure calcolando il numero di er-

rori commessi nell'assegnazione dell'etichetta categoriale. Inoltre, somministrando questo test alla fine di varie fasi di training è possibile anche rilevare eventuali cambiamenti delle rappresentazioni durante tutto l'apprendimento.

In sintesi, al di là del tipo di analisi che l'AFC task consente di eseguire rispetto ai classici compiti di classificazione (analisi che possono variare in base alla struttura delle categorie presentate durante la fase di training), questo compito di produzione permette di rilevare: a) il tipo di rappresentazione che i partecipanti si sono formati durante l'apprendimento (per es., esemplari, prototipi, regole) e, eventualmente, come cambiano durante le varie fasi dell'apprendimento; b) il tipo di elaborazione implementata durante l'osservazione degli esemplari nelle fasi di training (per es., attenzione alle singole caratteristiche o agli esempi nella loro globalità); l'accuratezza delle rappresentazioni (per es., errori nell'assegnazione delle etichette categoriali).

In conclusione, dunque, l'AFC task consente di osservare come una categoria viene appresa a partire dai suoi casi particolari, attraverso l'analisi di come questi casi vengono percepiti e rappresentati. Nel complesso, questo tipo di compito permette un'indagine "diretta" dei processi categoriali, cioè non mediata dagli stimoli di trasferimento e, quindi, di conseguenza, senza tutte le criticità che ne derivano.

6. Conclusioni

La categorizzazione, dunque, è quella complessa abilità cognitiva che parte dall'elaborazione percettiva, procede attraverso la rappresentazione delle informazioni e si manifesta nella classificazione.

Come abbiamo visto nella prima parte dell'elaborato, la maggior parte delle ricerche condotte in psicologia cognitiva si è concentrata principalmente sulla classificazione, che è un'importante funzione della categorizzazione ma non l'unica a dover essere indagata (Ross, Murphy 1999). Inoltre, si è visto come i compiti che sono stati tradizionalmente usati per studiare la classificazione soffrono di alcuni importanti limiti metodologici.

Il metodo sperimentale presentato nell'ultima parte di questo lavoro, l'Active Feature Composition task, è stato proposto con l'obiettivo di mostrare che indagini alternative sulla categorizzazione sono possibili e capaci di superare i limiti di cui soffrono i compiti di classificazione.

Nello studio in cui l'AFC task è stato implementato per la prima volta (Greco, Moretti 2017), ad esempio, si è trovata una forte relazione tra il tipo di elaborazione effettuata sugli stimoli del training e la qualità delle rappresentazioni categoriali raggiunte. Questo tipo di dati costituisce un tipo di informazione impossibile da ricavare con i tradizionali compiti di classificazione. Oltre al compito di produzione ai partecipanti è stato somministrato un test che richiedeva di valutare il livello di correttezza di una serie di regole presentate come possibili criteri per distinguere tra le categorie mostrate, e, successivamente, di esprimere verbalmente il criterio utilizzato per assegnare le etichette categoriali agli esemplari prodotti. L'analisi congiunta di questi diversi tipi di test ha permesso di indagare la relazione esistente tra il livello di consapevolezza circa il criterio utilizzato (e la sua accuratezza) e il tipo di elaborazione alla base delle rappresentazioni formatesi durante l'apprendimento.

L'AFC task, dunque, è un test di produzione che permette di rilevare quei processi coinvolti nell'acquisizione

e nella formazione delle categorie, senza i quali la classificazione di nuove informazioni non potrebbe avvenire.

Anche se questo paradigma sperimentale non permette di rendere conto in maniera esaustiva della categorizzazione nella sua complessità, l'obiettivo di questo lavoro è quello di incoraggiare le ricerche future a procedere in questa direzione. Un esempio di continuazione di questo primo studio effettuato con l'AFC task potrebbe essere, ad esempio, quello di integrare opportunamente compiti di produzione ed altre tipologie di test all'interno di un classico compito di classificazione, con l'ausilio di tecniche che permettono di misurare l'allocazione dell'attenzione o il carico della memoria di lavoro.

Considerando la varietà dei processi che intervengono nella categorizzazione, quali la percezione, l'attenzione, la memoria, fino al ragionamento e al giudizio, assumere un approccio innovativo e multidisciplinare sembra essere una condizione necessaria per una corretta comprensione del fenomeno, oltre che una soluzione perfettamente in linea con lo spirito delle scienze cognitive.

Riferimenti bibliografici

AHA D.W., KIBLER, D., ALBERT, M.K., (1991) *Instance-based learning algorithms*, «Machine Learning», 6(1):37–66.

ASHBY F.G., MADDOX W.T., (1998) *Stimulus categorization*, in Birnbaum M.H. (ed.), *Measurement, judgement, and decision making,* New York, Academic Press, pp. 251–301.

ASHBY, F.G., MADDOX, W.T., (2005) *Human category learning*, «Annual Review of Psychology», 56:149–178.

BROOKS, L.R., (1978) *Nonanalytic concept formation and memory for instances*, in Rosch E., Lloyd B.B., (ed.) *Cognition and categorization*, Hillsdale, NJ, Erlbaum.

BRUNER J.S., GOODNOW J.J., AUSTIN G.A., (1956) *A study of thinking*, Wiley, New York.

COVER T.M., HART P.E., (1967) *Nearest pattern classification*, «IEEE Transaction on Information Theory», 13(1):21–27.

DONKIN C., NEWELL B.R., KALISH M., DUNN J.C., NOSOFSKY R.M., (2015) *Identifying strategy use in category learning tasks: A case for more diagnostic data and models*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 41(4):933.

DUDA R., HART P., STORK D., (2001) *Pattern Classification*, 2nd ed., John Wiley & Sons, New York, NY.

ESTES, W.K., (1986) *Array models for category learning*, «Cognitive Psychology», 18:500–549.

ESTES W.K., (1994) *Classification and cognition*, New York, Oxford University Press.

GAGLIARDI F., (2009) *La categorizzazione tra psicologia cognitiva e machine learning: perché è necessario un approccio interdisciplinare*, «Sistemi Intelligenti», 21(3):489–502.

GAGLIARDI F., (2014) *La naturalizzazione dei concetti: aspetti computazionali e cognitivi,* «Sistemi Intelligenti», 26(2):283–298.

HITZMAN D.L., (1986) *Schema abstraction in a multiple–trace memory model*, «Psychological Review», 93:411–28.

GRECO A., MORETTI S., (2017) *Use of evidence in a categorization task: analytic and holistic processes*, «Cognitive Processing», 18(4):431–446.

HOFFMAN A.B., REHDER B., (2010) *The costs of super-*

vised classification: The effect of learning task on conceptual flexibility, «Journal of Experimental Psychology: General», 139(2):319.

HOMA D., STERLING S., TREPEL L., (1981) *Limitations of exemplar-based generalization and the abstraction of categorical information*, «Journal of Experimental Psychology: Human Learning and Memory», 7(6):418.

JOHANSEN M K., KRUSCHKE J.K., (2005) *Category representation for classification and feature inference*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 31(6):1433.

KRUSCHKE J.K., (1992) *ALCOVE: An exemplar–based connectionist model of category learning*, «Psychological Review», 99:22–44.

LAMBERTS K., (2000) *Information-accumulation theory of speeded categorization*, «Psychological Review», 107:227–60.

LANGLEY P., (1986) *On machine learning*, «Machine Learning», 1(1):5–10.

MARKMAN A.B., YAMAUCHI T., MAKIN V.S., (1997) *The creation of new concepts: A multifaceted approach to category learning*, in Ward T.B., Smith S.M., Vaid J., (ed.) *Creative thought: An investigation of conceptual structures and processes*, pp. 179–208.

MARKMAN A.B., ROSS B.H., (2003) *Category use and category learning*, «Psychological Bulletin», 129(4):592–613.

MEDIN D.L., SCHAFFER M.M., (1978) *Context theory of classification*, «Psychological Review», 85:207–238.

MICHIE D., SPIEGELHALTER D.J., TAYLOR, C.C., (1994) *Machine learning, neural and statistical classification*, Englewood Cliffs, NJ, Prentice Hall.

MORETTI S., GRECO A., (2016) *Costruire esempi per scoprire le rappresentazioni: un nuovo metodo d'indagine*

sulla categorizzazione, in Cruciani M., Gigliotta O., Marocco D., Miglino O., Moretti S., Ponticorvo M., Rubinacci F., (a cura di) *Apprendimento, cognizione e tecnologia*, Atti del convegno mid–term 2016 dell'Associazione Italiana di Scienze Cognitive (AISC), Università Studi di Napoli, pp. 142–149.

MURPHY G.L., (2002) *The big book of concepts*, Cambridge, MA, The MIT Press.

NILSSON H., OLSSON H., (2005), *Categorization vs. inference: Shift in attention or in representation?,* in Bara B.G., Barsalou L., Bucciarelli M., (a cura di) *Proceedings of the 27th Annual Conference of the Cognitive Science Society Stresa*, Italy, Cognitive Science Society, pp. 1642–1647.

NOSOFSKY R.M., (1986) *Attention, similarity and the identification–categorization relationship*, «Journal of Experimental Psychology: General», 115:39–57.

POSNER M.I., KEELE S.W., (1968) *On the genesis of abstract ideas,* «Journal of experimental psychology», 77(3p1):353.

POSNER M.I., KEELE S.W., (1970) *Retention of abstract ideas*, «Journal of Experimental Psychology», 83:304–308.

REED S.K., (1972) *Pattern recognition and categorization*, «Cognitive psychology», 3(3):382–407.

RICHARDSON D.C., KIRKHAM N.Z., (2004) *Multimodal events and moving locations: Eye movements of adults and 6–month–olds reveal dynamic spatial indexing*, «Journal of Experimental Psychology, General», 133:46–62.

RICHARDSON D.C., SPIVEY M.J., (2000) *Representation, space and Hollywood Squares: Looking at things that aren't there anymore*, «Cognition», 76:269–295.

ROSCH E., (1973) *Natural categories*, «Cognitive Psychol-

ogy», 4:328–50.

ROSCH E., (1975) *Cognitive reference points*, «Cognitive Psychology», 7:532–47.

ROSS B.H., MURPHY G.L., (1999) *Food for thought: Cross-classification and category organization in a complex real–world domain*, «Cognitive psychology», 38(4):495–553.

RUSSELL S.J., NORVIG P., (2002) *Artificial Intelligence. A Modern Approach*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ.

SCHOLZ A., VON HELVERSEN B., RIESKAMP J., (2015) *Eye movements reveal memory processes during similarity– and rule–based decision making*, «Cognition», 136:228-246.

SMITH J.D., MINDA J.P., (1998) *Prototypes in the mist: The early epochs of category learning*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 24(6):1411–1430.

SMITH J.D., MINDA J.P., (2001) *Journey to the center of the category: the dissociation in amnesia between categorization and recognition*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 27(4):984–1002.

TUNNEY R.J., FERNIE, G., (2012) *Episodic and prototype models of category learning*, «Cognitive processing», 13(1):41–54.

WHITTLESEA B.W., BROOKS L.R., WESTCOTT C., (1994) *After the learning is over: Factors controlling the selective application of general and particular knowledge*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 20(2):259.

WITTEN I.H., FRANK E., (2005) *Data mining practical machine learning tools and techniques with java implementations*, 2nd ed., Morgan Kaufmann, San Francisco.

YAMAUCHI T., MARKMAN A.B., (1998) *Category learning by inference and classification*, «Journal of Memory and Language», 39(1):124–148.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/978882551306611
pp. 211-234 (marzo 2018)

Rappresentazioni corticali

di ALESSIO PLEBE[1]

1. Introduzione

Per i neuroscienziati è pratica normale impiegare un vocabolario rappresentazionale nel caratterizzare diversi processi neurali, un uso che risulta particolarmente spontaneo quando i processi in esame riguardano le funzioni cognitive più pregiate. Mentre parlare di rappresentazioni per neuroni, i cui potenziali d'azione provocano una diretta contrazione di qualche muscolo, sembra ridondante, questo risulta del tutto naturale quando invece i neuroni in questione sono coinvolti nel pensare ad oggetti e fatti del mondo, magari nemmeno presenti, e meditarci sopra. Nonostante risulti, almeno in questi casi, del tutto ovvia, l'attribuzione di rappresentazioni a neuroni si è rivelata filosoficamente problematica e oggetto di non poche critiche, anche particolarmente aspre (Brooks, 1991; Chemero, 2000; Hutto, Myin, 2013; van Gelder, 1998).

Diversi dei guai sofferti dalla nozione di rappresentazione neurale sono direttamente ereditati da quelli, annosi, delle rappresentazioni mentali in senso ampio, uno dei temi più dibattuti della recente filosofia della mente (Ryder, 2009). In questo lavoro ci si limita ai termini che la questione ha assunto in ambito cognitivo, a partire dalla stretta connessione operata tra le nozioni di rappresentazione e di computazione, all'interno della

[1] Università di Messina. E–mail: aplebe@unime.it

classica teoria computazionale–rappresentazionale della mente. Si mostrerà anzitutto come, contrariamente a quanto spesso assunto, i supporti teorici al computazionalismo e al rappresentazionalismo sono ben lontani tra di loro. Ne segue che il poderoso e raffinato apparato matematico disponibile riguardo la computazione poco aiuta le rappresentazioni. Entrando tecnicamente nel merito, si sosterrà che l'attuale fondazione teorica delle rappresentazioni in scienze cognitive soffre di un vizio originale, per essere derivata dalla teoria matematica della misura, sviluppatesi dagli anni '70 ai '90 (Krantz, Luce, Suppes, Tversky, 1971; Luce, Krantz, Suppes, Tversky, 1990; Suppes, Krantz, Luce, Tversky, 1989). La finalità di tale teoria era la giustificazione di mappature operate tra sistemi empirici (oggetti e fenomeni fisici, biologici, sociali) e numeri. Il mondo empirico è notoriamente il regno di imprecisioni, vaghezza, ambiguità, fattori che hanno resa ardua la formalizzazione della misura, ma con il conforto che l'altro versante, quello dei numeri, eccelle oltre ogni altro in termini di rigore e sistematizzazione teorica. Vantaggio perso nel momento in cui, a partire da Swoyer (1991), si è presa a prestito la teoria della misura per le rappresentazioni: la mappatura ora diventa da mondo empirico a mondo mentale, ed è difficile dire quale dei due sia peggiore in termini di vaghezza e mancanza di definizione.

Volendo tentare di mantenere una nozione formale di rappresentazione, si è dell'avviso che un aiuto possa venire proprio dalla neuroscienza, iniziando a caratterizzare in un modo meno vago e liberale il codominio entro cui finisce la mappatura. Nel caso della misura era il comodo mondo dei numeri, poi diventato il ben poco definibile mondo mentale. Una via di mezzo può essere realizzata caratterizzando (in termini compatibili con la fondazione

matematica della misura) la forma matematica che assumono le rappresentazioni nei circuiti neurali. Pur essendo oggi ben lontani da una completa conoscenza di queste forme, è possibile formulare certi schemi di massima per determinati aggregati neurali, in particolare per ciò che concerne la corteccia cerebrale. Vi sono due punti a favore nel limitare lo studio a questa parte del cervello: da un lato è certamente la principale sede di rappresentazioni concettuali (Farhat, 2007; Fuster, 2008; Miller, Freedman, Wallis, 2002; Nieder, 2009; Noack, 2012); d'altro è la struttura cerebrale i cui ipotetici principi rappresentazionali godono di migliori analisi teoriche e ampi riscontri empirici.

2. Matrimonio tra stranieri

Si può senz'altro affermare che il matrimonio tra computazione e rappresentazioni sia stato felice, difficile trovare nelle scienze cognitive qualcosa che, finora, sia stato più fecondo dell'ampio ombrello che va sotto il nome di teoria computazionale-rappresentazionale della mente. Uno dei suoi principali fautori, Fodor (1981) ne ha voluto suggellare con forza l'unione, tramite il famoso slogan "no computation without representation". Recentemente sono stati sollevati dubbi, non tanto sulla serenità di questo matrimonio quanto sulla sua indissolubilità, sia Piccinini (2008) che Chalmers (2011) hanno mostrato come sia possibile concepire computazioni che non richiedono alcuna nozione di rappresentazione.

Quel che ci viene da osservare, e di solito viene trascurato, è che si tratta di un matrimonio tra stranieri. Computazione e rappresentazione hanno provenienze teoriche completamente distinte e diverse tra loro.

La prima vanta una storia prestigiosa, a cui si è agganciata l'accezione di computazione in scienze cognitive fin dall'inizio: la teoria tracciata dai padri fondatori della scienza informatica, Turing (1936) e Church (1941). Oltre a questo caposaldo, che rimane tuttora il riferimento principale, diversi dei progressi in informatica teorica hanno avuto importanti ricadute nello studio della mente, per esempio la teoria della trattabilità computazionale (Hartmanis, Stearns, 1965; Johnson, Papadimitriou, 1985; van Rooij, 2008). Notoriamente non sono mancati gli attacchi alla visione computazionale della mente, uno dei più celebri ed acuti dovuto a Searle (1990), così come le difese (Cordeschi, Frixione, 2007), a tutto vantaggio di un progresso nel definire cosa sia una computazione (Copeland, Posy, Shagrir, 2013; Scheutz, 2002). Negli ultimi anni si sono moltiplicate le proposte sempre più raffinate di idee teoriche di computazione che vadano bene sia per gli ordinari computer, che per il cervello di coloro che li costruiscono (e di altri animali) (Fresco, 2014; Miłkowski, 2013; Piccinini, 2015). Beninteso, questo progresso non ha certo messo a tacere le critiche al computazionalismo: un suo recente portavoce è, per esempio, Barrett (2016), ma non è questo il dibattito in cui intendiamo entrare, per una sua sintesi ben aggiornata rimandiamo a (Miłkowski, 2017).

Il coniuge che ci sta a cuore in questo matrimonio è rimasto molto in disparte dal fervore dialettico che vive il computazionalismo. Lo sforzo teorico riguardante la computazione ha sostanzialmente tenuto in disparte le rappresentazioni, che sono in genere considerate come qualcosa di scontato e che non richieda una propria fondazione. Sono rari, anche se notevoli (Rescorla, 2015), i tentativi di integrare teorie computazionali con un dettaglio formale su cosa siano le rappresentazioni, oltre a

non meglio esplicati simboli di un linguaggio mentale. Le ragioni risiedono proprio nelle distinte discendenze, e dopo aver sorvolato su quella che gode di miglior salute, il computazionalismo, nella prossima sezione si approfondirà la storia da cui derivano le rappresentazioni, e le conseguenze sul suo stato decisamente meno robusto.

3. Misure e mente

Fin dalla filosofia antica è stata accarezzata l'idea che le rappresentazioni mentali debbano poggiarsi su una qualche forma di *similarità* tra la rappresentazione stessa e ciò che è rappresentato, idea che assume forma compiuta in Hume (1748, cap. IX):

> All our reasoning concerning matter of fact are founded on a species of analogy, which leads us to expect from any cause the same events, which we have observed to result from similar causes. […] But where the objects have not so exact a similarity, the analogy is less perfect, and the inference is less conclusive; though still it has some force, in proportion to the degree of similarity and resemblance.

L'elaborazione di questa idea non aveva però potuto progredire molto, in quanto non era facile cercare di definire in modo compiuto in cosa consistesse una somiglianza o similarità parlando di rappresentazioni mentali. Affermare che la rappresentazione di faggio implichi una qualche similarità con i faggi è evidente: si allude ad una relazione profondamente diversa rispetto a quando si afferma, per esempio, che il faggio ha delle similarità con la quercia.

Il primo tentativo di formalizzare l'idea di similarità si deve a Russell (1927), che introdusse il *relation-number*

come elemento per stabilire che un insieme risulta rappresentativo dell'altro. È qualcosa del tutto simile al concetto meno tecnico di *struttura*, formalmente definito come la classe di tutte le relazioni che soddisfano una relazione di più alto livello, con una relazione data. Per Russell la similarità in termini di *relation–number* era un modo coerente per mettere in relazione il mondo fisico, con le sue relazioni, con un altro mondo in grado di coglierle, senza fare mistero su quale fosse il più privilegiato di questi ultimi mondi: quello delle percezioni sensoriali di un essere umano.

3.1 *Misure*

Negli anni '70 queste prime idee di Russell furono recepite per finalità molto diverse: gettare dei fondamenti sicuri su come dal mondo fisico si potesse transitare a quello tecnico-scientifico di una sua misurazione. Come dichiarano i fondatori della teoria della misura Krantz *et al.* (1971, p. XVII):

> Scattered about the literatures of economics, mathematics, philosophy, physics, psychology, and statistics are axiom systems and theorems that are intended to explain why some attributes of objects, substances, and events can reasonably be represented numerically. These results constitute the mathematical foundations of measurement.

La base che viene istituita come comune ad ogni genere di misura ricalca la strategia di Russell nel richiedere che sussista, tra il dominio rappresentato e quello della sua rappresentazione (codominio), una mappatura delle relazioni esistenti tra i due. Il genere di mappatura tra dominio e codominio è quella che matematicamente va sotto il nome di *omomorfismo*, che permette di mappare

due strutture algebriche, ovvero insiemi dotati di relazioni, preservando appunto tali relazioni.

Queste sono solamente le premesse, il lavoro serio poi richiede, per ogni tipo di misura: (1) di individuare nella porzione di mondo reale da rappresentare, che viene abitualmente chiamata *dominio empirico*; (2) quali relazioni sussistano; e (3) da esse dimostrare come teorema l'esistenza di un omomorfismo verso il codominio della misura. Non solo, occorre anche (4) dimostrare un teorema di unicità, vale a dire assicurarsi che la libertà di individuare omomorfismi non possa condurre a due tipi di misurazioni contraddittorie dello stesso dominio empirico. Questi sono i passi principali, ma non esauriscono il compito: occorre inoltre (6) preoccuparsi di altre faccende che rendano le misure praticabili, come l'analisi degli errori. Per vent'anni gli stessi autori (Luce *et al.*, 1990; Suppes *et al.*, 1989) hanno perseverato in questo minuzioso lavoro, coprendo un gran numero di misure in una trilogia che somma oltre 1500 pagine di densa matematica. In nessuna di queste 1500 pagine si trova la parola "mente", o un cenno ad essa.

L'opera di Krantz e collaboratori è stata possibile grazie ad un aspetto chiave: mentre il domino empirico è tipicamente qualcosa di sfumato, ambiguo, difficile da irreggimentare in leggi naturali, come tutte le cose del mondo che ci circonda, il codominio della sua rappresentazione sono i numeri, impalpabili entità che godono della più poderosa impresa razionale nel dotarli di strutture e proprietà, che si chiama matematica, di cui Krantz e collaboratori sono raffinati conoscitori.

3.2 *Mente*

Il prezioso vantaggio per la teoria della misura, di avere come elementi rappresentazionali i numeri, svanisce quando si è tentato di ritornare sui passi di Russell, che voleva impiantare un fondamento rigoroso all'idea di similarità che andasse bene sia per misurazioni tecniche della realtà fisica sia per le rappresentazioni mentali che scaturiscono dalla percezione di quella stessa realtà. Il primo ad impegnarsi nuovamente nell'impresa è Swoyer (1991), che adotta la strategia dell'omomorfismo ereditata dalla teoria della misura, nella declinazione che prende come dominio quello che lui chiama *intensional relational systems*. In termini discorsivi si tratta di un insieme di individui su cui sussistono relazioni (del primo ordine) e, in aggiunta, altre relazioni del secondo ordine che operano sulle relazioni del primo ordine. Un altro sistema dotato dello stesso genere di relazioni può essere detto una rappresentazione del primo se sussiste un omomorfismo tra i due. Swoyer caratterizza la sua nozione di rappresentazione come *strutturale*, pertanto in piena continuità con la prima formalizzazione di similarità in Russell, vista precedentemente. Fin qui siamo alle basi, prese dalla teoria della misura e specializzate agli *intensional relational systems*. Ma come visto precedentemente la parte corposa ed impegnativa della teoria della misura è trasformare la semplice enunciazione di un omomorfismo tra dominio empirico e sua rappresentazione in un teorema, e fornirne prima una dimostrazione di esistenza, e poi di unicità. Swoyer se ne è astenuto e nel suo lavoro ha presentato quattro esempi di come la sua formulazione di rappresentazione possa essere applicata, ma di questi tre non hanno nulla a che fare con la mente, e i codomini sono ancora una volta astrazioni matematiche. L'unico esempio che ri-

guarda le rappresentazioni mentali è solo accennato e meno sviluppato degli altri, e considera rappresentazioni di modelli alla (Johnson-Laird, 1983) e le rotazioni di (Shepard, Metzler, 1971). Swoyer è ben consapevole che una volta assunta la mente a codominio la faccenda diventa ardua, e tutto quel che si può fare è tentare di indicare una strada per chi voglia cimentarvisi (p.492):

> Three steps would have to be completed in order to develop this idea. First, […] we would have to adopt some set of axioms for ordinary geometrical relational systems […]. Second […] we would need to devise axioms for a Shepard IRS [(intensional relational systems)], which determine the structure of such relations as psychological coincidence or psychological rotation. Third, we would have to use these axioms to prove a representation theorem.

Nessuno ci si è mai cimentato.

3.3 *I guai per una mente come misura*

La concezione delle rappresentazioni mentali derivata dalla teoria della misura viene oggi considerata la più valida e rigorosa disponibile (Gallistel, King, 2010; Ramsey, 2007), purtuttavia è affetta da una serie di guai. Si è dell'avviso che si tratti di guai strettamente legati all'aver fatto uso di un apparato teorico che aveva ben altro come obiettivo: misurare, ovvero tradurre il mondo fisico in numeri. Il guaio più serio deriva dall'estremo liberalismo di cui gode la mappatura omomorfa: è sorprendentemente facile trovare sistemi che la soddisfano. Già qualcuno se ne era accorto riguardo ai *relation–number* di Russell, progenitori dell'omomorfismo, Newman (1928) dimostrò che è sufficiente per due insiemi avere la stessa cardinalità per escogitare *relation-number* che mettano d'accordo

qualunque loro relazione. Peggio ancora: McLendon (1955) fece vedere che anche prendendo due insiemi di cardinalità diversa, è possibile mapparne sempre qualche loro insieme di partizioni. Non si pensi che questi duri colpi alla formalizzazione dell'idea di similarità di Russell abbiano avuto ripercussioni sulla teoria della misura, perché come visto non solo l'omomorfismo deve venir dimostrato, ma anche la sua unicità.

L'ubiquità delle rappresentazioni definite tramite omomorfismo diventa però imbarazzante quando la stessa definizione la si vorrebbe capace di caratterizzare le rappresentazioni mentali. Dei perfetti omomorfismi possono essere instaurati tra il serbatoio di carburante di una macchina e il suo indicatore nel cruscotto, ma pochi gli attribuirebbero lo statuto di un sistema di rappresentazione alla stessa stregua della mente. Che sia una legittima rappresentazione nella stessa accezione matematica è piuttosto ovvio: si tratta proprio di uno strumento misuratore. Sono stati proposti svariati tentativi di rimedio (Bartels, 2006; Gallistel, King, 2010; Isaac, 2013; Shea, 2014): uno di quelli maggiormente ricorrenti consiste nel richiedere alle rappresentazioni di svolgere un qualche ruolo per l'organismo che le possiede, di essere in grado di avere un impatto sul sistema motorio, causando azioni diverse a seconda del contenuto corrente delle rappresentazioni. Ai rimedi facilmente seguono i controesempi. I ritmi circadiani sono meccanismi biochimici che producono segnali endogeni in fase con il ciclo solare. Si ritrovano non solo negli animali, ma anche nelle piante, e in alcune di esse hanno un impatto su azioni, come quella di orientare le foglie (Kay, 1997). Pur se a un livello superiore rispetto all'indicatore del carburante, anche a questo meccanismo pochi attribuirebbero lo statuto di rappresentazione mentale. Le discussioni sui guai della mente come misura e sulle

strategie per una sua difesa – come quelle appena descritte – abbondano (Morgan, 2014), ma ogni strategia assume formato discorsivo, ovvero non riesce a tradursi in un rafforzamento dell'impianto formale delle rappresentazioni basato sulla teoria della misura.

4. Strutture familiari alla corteccia

Si è dell'idea che oggi si potrebbero percorrere strade ben diverse, e tentare di agire proprio sulla parte debole del modello di omomorfismo assunto come base delle rappresentazioni strutturali. Quando il modello forniva il teorema principale di misura funzionava benissimo, perché il suo codominio erano i numeri, che consente ampi e ben controllabili margini di manovra. I guai come abbiamo visto iniziano quando al suo posto ci finisce la mente: difficile trovare qualcosa di più indefinito e vago. Un rimedio che si ritiene oggi sia diventato possibile è quello di lavorare sul codominio impiegando le conoscenze attuali su come vanno a strutturarsi matematicamente le informazioni nel cervello, più specificatamente nella sua parte che è sede privilegiata di rappresentazioni mentali: la corteccia cerebrale.

Vi sono diverse strade che potrebbero essere intraprese per il progetto qui perorato, di fondare le rappresentazioni neurali su una schematizzazione matematica plausibile del codominio della relazione di misura. Vi sono diverse caratterizzazioni note e ampiamente percorse della corteccia cerebrale, quali l'organizzazione a mappatura topologica sulla superficie a due dimensioni, ortogonale a quella verticale che caratterizza invece le colonne corticali (Felleman, Van Essen, 1991; Hubel, Wiesel, 1959; Krubitzer, 1995; Mountcastle, 1957). Pur se si tratta di

principi estremamente fecondi e ben sviluppati in determinate aree corticali, come quella visiva (Hubel, Wiesel, 1968; Tootell, Silverman, Hamilton, Switkes, De Valois, 1988; Wiesel, Hubel, 1965), l'idea che vi sia un principio sottostante generalizzabile all'intera corteccia cerebrale è stato recentemente messo in seria discussione (Horton, Adams, 2005; Maçarico da Costa, Martin, 2010).

Una strada che – pur non consentendo livelli di dettagli matematici paragonabili alla mappatura topologica (Bednar, Wilson, 2015; Stevens, Law, Antolik, Bednar, 2013) – presenta il vantaggio di una maggior generalizzazione è la codifica a popolazione neurale. Si tratta di un concetto suggerito a più riprese, con denominazioni e sottigliezze definitorie diverse: da "codifica distribuita" (Hinton, McClelland, Rumelhart, 1986), a "spazi neurali vettoriali" (P. M. Churchland, 1989; P. S. Churchland, Sejnowski, 1994; Sejnowski, 1998), infine a "codifica a popolazione" (Quian Quiroga, Panzeri, 2013; Zemel, Dayan, Pouget, 1998). L'idea comune è che la corteccia codifichi un certo dominio categoriale attraverso la combinazione dell'attività di molti neuroni in una stessa popolazione, per alcuni come Yuste (2015) ritenuto uno dei paradigmi portanti nelle attuali neuroscienze.

Una delle prime evidenze empiriche della codifica a popolazione nella corteccia è dovuta a (Georgopoulos, Schwartz, Kettner, 1986) per la rappresentazione di direzioni di movimento nella corteccia motoria; successivamente la parte dominante degli esperimenti ha riguardato categorizzazione di oggetti visivi, con molte conferme del concorso combinatorio di molti neuroni nel codificare oggetti e stimoli visivi (Abbott, Rolls, Tovee, 1996; Pasupathy, Connor, 2002; Rolls, Tovee, 1995; Sakai, Naya, Miyashita, 1994). Ma non sono mancate conferme per al-

tre aree corticali, sia percettive, come nella codifica a popolazione di localizzazioni sonore (Fitzpatrick, Batra, Stanford, Kuwada, 1997) oppure olfattive (Miura, Mainen, Uchida, 2012), che amodali, per esempio gli studi di Stokes *et al.* (2013) sulla codifica di decisioni comportamentali nella corteccia prefrontale, e di Chikazoe, Lee, Kriegeskorte, and Anderson (2014) sulla valenza affettiva nella corteccia orbitofrontale.

4.1 *Formalizzazione della codifica a popolazione*

Non è questa la sede opportuna per una esposizione matematica della proposta di formalizzazione della codifica a popolazione. Ne verrà invece fornita una sintesi descrittiva, utile ai fini qui perseguiti, di mostrare come possa riempire quel vuoto denunciato nell'ereditare le rappresentazioni strutturali dalla teoria della misura, con codominio la mente. Un trattamento più rigoroso di alcuni aspetti della presente proposta si possono trovare in (Plebe, De La Cruz, 2016).

Supponendo di prendere in esame un certo dominio empirico, costituito da oggetti concreti del mondo, che consideriamo raggruppati in un certo numero di categorie distinte (per esempio in osservanza delle norme semantiche di un linguaggio naturale), è possibile definire funzioni che "misurino" l'attivazione di neuroni in una popolazione influenzata in qualche modo dalla percezione di stimoli in cui siano presenti oggetti di questo dominio. La misura dell'attivazione può essere sopperita in modo convenzionale con una delle modalità tipiche di caratterizzare attività neurale, per esempio mediante frequenza di potenziali d'azione in finestre temporali prefissate (de la Rocha, Doiron, Shea–Brown, Josić, Reyes, 2007).

Il passaggio chiave del procedimento proposto consiste nell'istituire un ordinamento di tutti i neuroni coinvolti nella rappresentazione, separato per ogni categoria. Questo può essere effettuato in modo oggettivo costruendo, per ogni neurone e per ogni categoria, due distribuzioni di valori: una di tutte le risposte a stimoli che ricadano nella categoria in esame, e un'altra con le risposte a stimoli di tutte le categorie diverse da quella in esame. Il parametro scalare su cui costruire l'ordinamento non è altro che la significatività statistica della differenza tra queste due distribuzioni. Una volta ottenuti questi ordinamenti per ogni categoria, si possono istituire dei *codici a popolazione*, prendendo per ogni categoria un numero adeguato dei primi neuroni più rispondenti a quella categoria. Il passo successivo è definire la *classificazione* concettuale di uno stimolo come il confronto tra le effettive attività prodotte in risposta nella popolazione di neuroni e i codici a popolazione per le diverse categorie: si considera che lo stimolo sia classificato nella categoria il cui codice assomiglia maggiormente alla distribuzione di attivazioni.

Si può mostrare come questo procedimento si presti ad una mappatura omomorfa tra il dominio empirico e quello corticale. Per farlo occorre istituire qualche ulteriore formalismo che catturi i criteri su cui si fondano le categorie nel dominio empirico. Uno strumento adeguato allo scopo è l'analisi concettuale formale di Ganter and Wille (1999), ampiamente adottata nella scienza informatica. Pur consapevoli della banalizzazione possiamo sintetizzarla semplicemente come l'adozione della vetusta definizione dei concetti sulla base di proprietà necessarie e sufficienti (Smoke, 1932), vituperata da Rosch (1978) in poi in psicologia, ma più che buona per gli informatici. Ottima nel nostro caso, dove non ha nulla a che vedere con la concettua-

lizzazione mentale, è semplicemente una strategia per rendere in termini matematici un certo criterio con cui un domino empirico può essere categorizzato. Operando questa mossa, non è difficile espandere dal lato mentale la formulazione delle attivazioni della popolazione di neuroni che rispondono a stimoli del dominio empirico. Tali attivazioni saranno infatti causate da afferenze da altre aree corticali, o dal talamo, descrivibili pertanto in una gerarchia che, al livello più basso, riguarderà proprio la percezione delle proprietà su cui si basa la categorizzazione empirica. Esempi di questa espansione matematica per proprietà che riguardano forma e colore di oggetti visivi sono in (Plebe, De La Cruz, 2016).

4.2 *Virtù (e vizi) della formalizzazione proposta*

È possibile mostrare, anche se solo in modo intuitivo, come la strutturazione mediante codifica a popolazione presenti più di una virtù, anzitutto risponde naturalmente a quel requisito aggiuntivo che in diversi hanno proposto per rimediare all'ubiquità delle rappresentazioni fondata su omomorfismi, ovvero la capacità di esercitare un ruolo sul sistema motorio. Proprio una delle prime evidenze empiriche della codifica a popolazioni (Georgopoulos *et al.*, 1986) riguardava la rappresentazione dello spazio dei movimenti delle braccia nella corteccia motoria. In un modello neurocomputazionale di decisioni morali (Plebe, 2015) facente uso di un formalismo di rappresentazioni corticali compatibile con quello qui descritto, le diverse codifiche a popolazione di oggetti visivi e del loro sapore combinati nella corteccia orbitofrontale aveva effetto causale sulle attivazioni nella corteccia prefrontale ventro-mediale, producendo scelte di azioni diverse. In questo caso la rappresentazione non era già di per sé motoria, ma veniva impie-

gata in altri sistemi corticali che alla fine si riflettono in comandi motori.

Non è difficile pensare a come il quadro rappresentazionale proposto si presti a una modalità di utilizzo delle rappresentazioni particolarmente importante in senso mentale, non solamente nel dirigere azioni, ma anche per ragionare su rappresentazioni. Supponiamo che il dominio empirico di interesse riguardi il mondo vegetale, e che si siano instaurate due codifiche a popolazione corrispondenti alle categorie faggio e quercia. Il ragionare su cosa accomuni questi due alberi nel mondo reale diventa possibile tramite l'intersezione degli insiemi dei neuroni nella popolazione che concorrono alla codifica di faggio e quercia, e quindi la sovrapposizione delle loro afferenze, veicolanti proprietà comuni. Viceversa riflettere sulle differenze tra faggio e quercia è possibile dall'insieme–differenza tra gli insiemi dei neuroni nella popolazione che concorrono alla codifica dei due alberi, e, nuovamente, alla provenienza delle loro afferenze.

Alcune avvertenze sono d'obbligo riguardo ad alcune disinvolture compiute nel delineare il formalismo di rappresentazione corticale che ne viziano la correttezza. Nel mostrare come la codifica a popolazione instauri effettivamente un omomorfismo abbiamo considerato che la categorizzazione del dominio empirico risponda a criteri di proprietà necessarie e sufficienti, e che a livello della corteccia, i neuroni che rispondono agli stimoli della categoria abbiano afferenze che veicolano precisamente quelle proprietà. Si tratta di una grossolana semplificazione. I generi di proprietà a cui si fa riferimento parlando di categorie (colore, altezza, larghezza, forma, peso specifico, tessitura superficiale, e così via) non hanno nessuna corrispondenza esclusiva con proiezioni corticali. Il livello di integrazione e olismo del cervello è elevato, e anche aree pri-

marie una volta ritenute unimodali lo sono solamente in prima approssimazione (Klemen, Chambers, 2012). Inoltre nell'usare l'analisi concettuale formale di (Ganter, Wille, 1999) si è sorvolato su una differenza non da poco rispetto agli elementi base degli insiemi presi in considerazione, poiché, mentre in quelli del modello formale del dominio empirico tali elementi sono individui, la loro popolazione di risposte corticali non è in relazione ad un individuo, bensì ad una tra le infinite sue modalità di presentarsi al sistema sensoriale. Si tratta di vizi che non inficiano l'impianto in linea di principio, ma la matematica agile di cui il progetto di formalizzazione si è finora dotato ha il prezzo di semplificazioni non indifferenti rispetto a come vanno davvero le cose nel cervello, e voler essere più realisti rispetto ai due vizi appena descritti potrebbe complicare non di poco la matematica necessaria.

5. Conclusioni

Si è voluto mostrare come le rappresentazioni mentali, nella loro miglior formalizzazione disponibile, derivata dalla teoria della misura, non siano messe troppo bene, scricchiolando di fronte a serie obiezioni, che si ritiene siano proprio conseguenze della loro derivazione. Ancor più precisamente, conseguenza della sostituzione del comodo codominio dei numeri, nella teoria della misura, con l'oscuro e scivoloso codominio della mente stessa. Si è anche proposto un rimedio, oggi più che ragionevole: tentare di delineare matematicamente il codominio mentale sfruttando quel che oggi è noto su come codificano i sistemi neurali, in particolare la corteccia.

Riferimenti bibliografici

ABBOTT, L. F., ROLLS, E., TOVEE, M. J. (1996) Representational capacity of face coding in monkeys. *Cerebral Cortex*, *6*, 498–505.

BARRETT, L. (2016) Why brains are not computers, why behaviorism is not satanism, and why dolphins are not aquatic apes. *The Behavior Analyst*, *39*, 9–23.

BARTELS, A. (2006) Defending the structural concept of representation. *Theoria*, *55*, 7–19.

BEDNAR, J. A., WILSON, S. P. (2015) Cortical maps. *The Neuroscientist*, *1073858415597645*, 1–14.

BROOKS, R. A. (1991) Intelligence without representation. In J. Haugeland (Ed.), *Mind design ii* (pp. 395–420) Cambridge (MA): MIT Press. (Second edition, 1997)

CHALMERS, D. (2011) A computational foundation for the study of cognition. *Journal of Consciousness Studies*, *12*, 323-357.

CHEMERO, A. (2000) Anti-representationalism and the dynamical stance. *Philosophy of Science*, *67*, 625–647.

CHIKAZOE, J., LEE, D. H., KRIEGESKORT, N., ANDERSON, A. K. (2014) Population coding of affect across stimuli, modalities and individuals. *Nature Neuroscience*, *17*, 1114–1122.

CHURCH, A. (1941) *The calculi of lambda conversion*. Princeton (NJ): Princeton University Press.

CHURCHLAND, P. M. (1989) *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge (MA): MIT Press.

CHURCHLAND, P. S., SEJNOWSKI, T. (1994) *The computational brain*. Cambridge (MA): MIT Press.

COPELAND, J. B., POSY, C. J., SHAGRIR, O. (Eds.) (2013) *Computability: Turing, gödel, church, and beyond*. Cambridge (MA): MIT Press.

CORDESCHI, R., FRIXIONE, M. (2007) Computationalism under attack. In M. Marraffa, M. D. Caro, F. Ferretti (Eds.) *Cartographies of the mind* (pp. 37–49) Berlin: Springer-Verlag.

DE LA ROCHA, J., DOIRON, B., SHEA-BROWN, E., JOSIĆ, K., REYES, A. (2007) Correlation between neural spike trains increases with firing rate. *Nature*, *448*, 802–809.

FARHAT, N. H. (2007) Corticonic models of brain mechanisms underlying cognition and intelligence. *Physics of Life Reviews*, *4*, 223–252.

FELLEMAN, D. J., VAN ESSEN, D. C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

FITZPATRICK, D. C., BATRA, R., STANFORD, T. R., KUWADA, S. (1997) A neuronal population code for sound localization. *Nature*, *388*, 871–874.

FODOR, J. (1981) *Representations: Philosofical essay on the foundation of cognitive science*. Cambridge (MA): MIT Press.

FRESCO, N. (2014) *Physical computation and cognitive science*. Berlin: Springer-Verlag.

FUSTER, J. M. (2008) *The prefrontal cortex*. New York: Academic Press. (fourth edition)

GALLISTEL, C. R., KING, A. P. (2010) *Memory and the computational brain: Why cognitive science will transform neuroscience*. New York: John Wiley.

GANTER, B., WILLE, R. (1999) *Formal concept analysis: mathematical foundations*. Berlin: Springer-Verlag.

GEORGOPOULOS, A. P., SCHWARTZ, A. B., KETTNER, R. E. (1986) Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

HARTMANIS, J., STEARNS, R. E. (1965) On the computational complexity of algorithms. *Transaction of American Mathematical Society*, *117*, 285–306.

HINTON, G. E., MCCLELLAND, J. L., RUMELHART, D. E. (1986) Distributed representations. In D. E. Rumelhart, J. L. McClelland (Eds.) *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 77–109) Cambridge (MA): MIT Press.

HORTON, J. C., ADAMS, D. L. (2005) The cortical column: a structure without a function. *Philosophical transactions of the Royal Society B*, *360*, 837–862.

HUBEL, D., WIESEL, T. (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.

HUBEL, D., WIESEL, T. (1968) Receptive fields and functional architecture of mokey striate cortex. *Journal of Physiology*, *195*, 215–243.

HUME, D. (1748) *An enquiry concerning human understanding*. London: A. Millar.

HUTTO, D. D., MYIN, E. (2013) *Radicalizing enactivism: basic minds without content*. Cambridge (MA): MIT Press.

ISAAC, A. M. (2013) Objective similarity and mental representation. *Australasian Journal of Philosophy*, *91*, 683–704.

JOHNSON, D. S., PAPADIMITRIOU, C. H. (1985) Computational complexity. In E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, D. B. Shmoys (Eds.) *The travelling salesman problem: a guided tour of combinatorial optimization* (pp. 37–85) New York: John Wiley.

JOHNSON-LAIRD, P. (1983) *Mental models: towards a cognitive science of language, inference and consciousness*. Cambridge (UK): Cambridge University Press.

KAY, S. A. (1997) PAS, present, and future: clues to the origins of circadian clocks. *Science*, *276*, 753–754.

KLEMEN, J., CHAMBERS, C. D. (2012) Current perspectives and methods in studying neural mechanisms of multi-

sensory interactions. *Neuroscience and Biobehavioral Reviews*, *36*, 111–133.

KRANTZ, D., LUCE, D., SUPPES, P., TVERSKY, A. (1971) *Foundations of measurement – volume i additive and polynomial representations*. New York: Academic Press.

KRUBITZER, L. (1995) The organization of neocortex in mammals: are species differences really so different? *Trends in Neuroscience*, *8*, 408–417.

LUCE, D., KRANTZ, D., SUPPES, P., TVERSKY, A. (1990) *Foundations of measurement – volume iii representation, axiomatization, and invariance*. New York: Academic Press.

MAÇARICO DA COSTA, N., MARTIN, K. A. C. (2010) Whose cortical column would that be? *Frontiers in Neuroanatomy*, *4*, 16.

MCLENDON, H. J. (1955) Uses of similarity of structure in contemporary philosophy. *Mind*, *64*, 79–95.

MIŁKOWSKI, M. (2013) *Explaining the computational mind*. Cambridge (MA): MIT Press.

MIŁKOWSKI, M. (2017) Objections to computationalism. a short survey. In *Annual meeting of the cognitive science society* (pp. 2723–2728)

MILLER, E. K., FREEDMAN, D. J., WALLIS, J. D. (2002) The prefrontal cortex: Categories, concepts and cognition. *Philosophical Transactions: Biological Sciences*, *357*, 1123–1136.

MIURA, K., MAINEN, Z. F., UCHIDA, N. (2012) Odor representations in olfactory cortex: Distributed rate coding and decorrelated population activity. *Neuron*, *74*, 1087–1098.

MORGAN, A. (2014) Representations gone mental. *Synthese*, *191*, 213-244.

MOUNTCASTLE, V. (1957) Modality and topographic

properties of single neurons in cats somatic sensory cortex. *Journal of Neurophysiology*, *20*, 408–434.

NEWMAN, M. H. A. (1928) Mr. Russell's "causal theory of perception". *Mind*, *37*, 137–148.

NIEDER, A. (2009) Prefrontal cortex and the evolution of symbolic reference. *Current Opinion in Neurobiology*, *19*, 99–108.

NOACK, R. A. (2012) Solving the "human problem": The frontal feedback model. *Consciousness and Cognition*, *21*, 1043-1067.

PASUPATHY, A., CONNOR, C. E. (2002) Population coding of shape in area v4. *Nature Neuroscience*, *5*, 1332–1338.

PICCININI, G. (2008) Computation without representation. *Philosophical studies*, *137*, 205–241.

PICCININI, G. (2015) *Physical computation: A mechanistic account*. Oxford (UK): Oxford University Press.

QUIAN QUIROGA, R., PANZERI, S. (Eds.) (2013) *Principles of neural coding*. Boca Raton (FL): CRC Press.

PLEBE, A. (2015) Neurocomputational model of moral behaviour. *Biological Cybernetics, 109,* 685-699.

PLEBE, A., DE LA CRUZ, V. M. (2016) *Neurosemantics – Neural Processes and the Construction of Linguistic Meaning*. Berlin: Springer.

RAMSEY, W. M. (2007) *Representation reconsidered*. Cambridge (UK): Cambridge University Press.

RESCORLA, M. (2015) The representational foundations of computation. *Philosophia Mathematica*, *23*, 338–366.

ROLLS, E., TOVEE, M. J. (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*, 713–726.

ROSCH, E. (1978) Principles of categorization. In E. Rosch, B. Lloyd (Eds.) *Cognition and categorization.* Mahwah (NJ): Lawrence Erlbaum Associates.

RUSSELL, B. (1927) *The analysis of matter*. London: Harcourt.

RYDER, D. (2009) Problems of representation I: nature and role. In J. Symons, P. Calvo (Eds.) *The routledge companion to philosophy of psychology* (pp. 233–250) London: Routledge.

SAKAI, K., NAYA, Y., MIYASHITA, Y. (1994) Neuronal tuning and associative mechanisms in form representation. *Learning and Menory*, *1*, 83–105.

SCHEUTZ, M. (Ed.) (2002) *Computationalism – new directions*. Cambridge (MA): MIT Press.

SEARLE, J. R. (1990) Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association*, *64*, 21–37.

SEJNOWSKI, T. J. (1998) Neural populations revealed. *Nature*, *332*, 308.

SHEA, N. (2014) Exploitable isomorphism and structural representation. *Proceedings of the Aristotelian Society*, *114*, 123–144.

SHEPARD, R. N., METZLER, J. (1971) Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

SMOKE, K. L. (1932) An objective study of concept formation. *Psychological Monographs*, *42*, 1–46.

STEVENS, J.-L. R., LAW, J. S., ANTOLIK, J., BEDNAR, J. A. (2013) Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *JNS*, *33*, 15747–15766.

STOKES, M. G., KUSUNOKI, M., SIGALA, N., NILI, H., GAFFAN, D., DUNCAN, J. (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*, 364–375.

SUPPES, P., KRANTZ, D., LUCE, D., TVERSKY, A. (1989) *Foundations of measurement – volume ii geometrical, threshold, and probabilistic representations*. New York: Academic Press.

SWOYER, C. (1991) Structural representation and surrogative reasoning. *Synthese*, *87*, 449–508.

TOOTELL, R. B., SILVERMAN, M. S., HAMILTON, S. L., SWITKES, E., DE VALOIS, R. (1988) Functional anatomy of the macaque striate cortex. V. spatial frequency. *Journal of Neuroscience*, *8*, 1610–1624.

TURING, A. (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *42*, 230–265.

VAN GELDER, T. (1998) The dynamical hypothesis in cognitive science. *Behavioral and Brain Science*, *21*, 615–665.

VAN ROOIJ, I. (2008) The tractable cognition thesis. *Cognitive Science*, *32*, 939-984.

WIESEL, T., HUBEL, D. (1965) Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, *28*, 1041–1059.

YUSTE, R. (2015) From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, *16*, 1–11.

ZEMEL, R. S., DAYAN, P., POUGET, A. (1998) Probabilistic interpretation of population codes. *Neural Computation*, *10*, 403–430.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/978882551306612
pp. 235-264 (marzo 2018)

Rappresentazione e *embodiment* debole

L'intreccio tra astrazione e esperienza e l'ipotesi di una doppia elaborazione multimodale e amodale dei concetti

di ANDREA VELARDI[1]

1. Sistema sensorimotorio e teoria dei concetti

Il dibattito sulla relazione tra astratto e concreto anima le scienze cognitive divise tra una nozione di rappresentazione amodale, astratta, proposizionale e arbitraria e l'approccio *embodied* fondato sulle nozioni di simulazione (che non vuol dire fare, ma lega il *processing* dei concetti alle *perfomance* pratiche e all'esecuzione di schemi motori che noi compiamo con gli oggetti evocati dai concetti), di *riattivazione* del sistema sensorimotorio degli schemi di interazioni avute in passato con gli oggetti che sono richiamati dalle parole e dalle frasi in cui compaiono i concetti; di simbolo percettivo e multimodale (Barsalou, 1999, 2008, 2010; Borghi, Binkofski 2014; Gallese, Lakoff, 2005[2]; per una rassegna recente Galetzka 2017).

L'*embodiment* ha sminuito il ruolo delle rappresentazioni a favore di un approccio enattivista e *bodily grounded* radicale (Gallagher 2017). Sembra però che il suo vero bersaglio sia una rappresentazione di tipo fodoriano

[1] Università di Messina, Dipartimento di Scienze Cognitive, Psicologiche, Pedagogiche e degli Studi Culturali COSPECS. E–mail: velardi.velardi@gmail.com

[2] Questi autori pensano comunque che i concetti astratti possano essere suscettibili di rappresentazioni amodali.

classico, non rendendo giustizia ai tentativi filosofici e cognitivi di rispondere al problema del *grounding* e legando percezione e processo della generazione delle idee. Fodor (1976) è autore della "teoria computazionale e rappresentazionale della mente" secondo cui il pensiero è costituito di simboli astratti tipici di una rete computazionale il cui ruolo causale è definito dalle regole di un *linguaggio del pensiero* innato. La mente è un elaboratore elettronico capace di manipolare questi simboli solo sintatticamente, senza alcuna finestra sul loro contenuto semantico, come un sistema di diagnosi mediche riesce a produrre il suo referto senza comprendere nulla di cosa sono i sintomi e le malattie[3]. In quest'ottica i concetti sono pensati come qualcosa di fortemente amodale, tutta la semantica è riducibile a regole sintattiche e tutta la nostra conoscenza è traducibile in simboli di carattere proposizionale[4]. Questo modello ha caratterizzato tutta la scienza cognitiva classica con una forte egemonia, attestata dal forte dibattito sul formato dei concetti che ha visto opporsi i proposizionalisti (Pylyshyn 1973, 1984, 2004[5]) e i pittorialisti (Kosslyn 1980, 1983; Kosslyn *et al.* 2006).

I primi negano che le immagini mentali possano essere candidate per la manipolazione computazionale concettuale, affermando che esse sono semplicemente un epifeno-

[3] Si ricorderà che questa tesi si oppone alla confutazione paradigmatica dell'argomento della *stanza cinese* di SEARLE (1980), in cui si dimostra che la semantica non è equivalente alla sintassi.

[4] In questa sede tralasciamo di entrare nel dettaglio della teoria dei concetti di Fodor che ha subito notevoli variazioni negli anni, mantenendo però il forte carattere proposizionale e amodale. Per una breve analisi della sua teoria dei concetti si veda il secondo contributo all'interno di questo volume dedicato alla relazione fra concetti e selezione lessicale.

[5] L'autore, nella sua *Visual Indexing Theory*, prevede solo un meccanismo pre–concettuale e pre–attentivo per il *processing* degli elementi visivi salienti di un contesto, attraverso cui vengono anche fornite le localizzazioni spazio–temporali degli oggetti.

meno analogico–sensoriale emergente da strutture che restano fortemente computazionali e proposizionali, allo stesso modo in cui le immagini dello *screen* di un computer sono frutto della codificazione sintattica dei *byte* e il loro contenuto visivo (come quello semantico dei concetti nel linguaggio del pensiero di Fodor) non ha uno statuto cognitivo autonomo, né un ruolo funzionale all'interno del *processing* essendo solo frutto di manipolazione di regole e codici computazionali astratti; i secondi dimostrano invece, anche con dei sofisticati esperimenti, che le immagini mentali, soprattutto quelle visive, sono un analogato sensoriale, coerente e insostituibile, della realtà e si prestano a operazioni di natura concettuale–computazione svolgendo quindi un ruolo funzionale vero e proprio all'interno dei processi cognitivi, senza il quale il soggetto non può orientarsi all'interno della realtà.

Questo dibattito è più vicino alle dispute filosofiche sui concetti, che sembrano essere state più aperte a discutere il contributo delle immagini mentali nella generazione delle idee, e anche agli sviluppi della teoria della categorizzazione nella scienza cognitiva precedente alla svolta dell'*embodiment*, come la teoria dei prototipi e degli esemplari (Murphy 2002; Rosch *et al*., 1976; Velardi 2005). Dal punto di vista filosofico, si pensi alla discussione del problema della generalizzazione e dell'astrazione in Locke, Berkeley, Hume e in generale in tutta la storia della filosofia (Velardi 2013). L'intreccio tra idea generale astratta, rappresentazione percettiva e oggetto particolare è centrale nelle teorie della conoscenza umana. I concetti non nascono da un'ascesa amodale e astratta che spoglia gli oggetti delle loro caratteristiche fisiche e sensoriali, ma mantengono una relazione con l'esperienza multimodale. Ne *La vita delle idee* si discute il problema dell'astrazione e il problema della generalizzazione distinguendoli e in-

trecciandoli fra di loro (Velardi 2013). Essi sono centrali nella discussione dell'*embodiment*.

Il *problema della generalità* riguarda il fatto che ogni immagine rappresenta un oggetto particolare, ma non si possono concepire idee generali a partire da oggetti e nomi particolari, perché il nome che indica un oggetto generale enuncia le caratteristiche particolari dell'oggetto individuale cui si riferisce: la parola *cane* si riferisce ad un *cane* particolare che non posso generalizzare come tale. Il *problema dell'astrazione* riguarda l'impossibilità di derivare una conoscenza generale dall'esperienza singola in quanto essa è sempre esperienza di più cose particolari. Come vedremo (§4) i due problemi sono centrali nella critica all'*emdobiment* radicale del filosofo Guy Dove (2016), fautore di una teoria ibrida della rappresentazione multipla *embodied* e amodale dei concetti, che pensa che il problema della *generalization* insieme a quello del *disembodiment* differenziato, sia il problema più spinoso per i concetti concreti, per i concetti sovraordinati (ANIMALE, FRUTTA, MOBILE) e soprattutto per i concetti astratti come *libertà*, *giustizia*, *verità*, *democrazia*, *numeri dispari*, *elettroni*, non dotati di referenti singoli e tangibili. D'altra parte la generalizzazione sarebbe risolta meglio da una teoria rappresentazionale amodale che non da una teoria multimodale e sensorimotoria dei concetti.

I due problemi della generalizzazione e dell'astrazione partono dalla premessa sbagliata che non si possa avere generalità a partire dall'individualità degli oggetti e che l'immagine mentale non possa avere una qualità astratta, cosa invece confermata dalla teoria del *livello di base* dell'astrazione (Rosch *et al*., 1976) costituito da categorie di oggetti di base (*cane* e *tigre* per il sovraordinato ANIMALE, *mela* e *mirtillo* per il sovraordinato FRUTTA, *sedia* e *divano* per il sovraordinato MOBILE) che possono

fornire immagini mentali visive alla categoria permettendo un maggiore aggancio con gli esemplari concreti della esperienza sensoriale e motoria. Prospettiva questa ribadita dallo sviluppo della teoria dei prototipi nella teoria degli esemplari e nella teoria della teoria che ha approfondito ulteriormente il rapporto con i referenti individuali e concreti della realtà (Murphy 2002, Velardi 2005).

I concetti astratti come *libertà* e *giustizia* erano stati esclusi dalla teoria dei prototipi perché la Rosch ne rilevava la mancanza di prototipicità e il difficile trattamento sperimentale (Velardi 2005). Essi costituiscono la sfida più straordinaria per il resoconto *embodied* e *grounded*.

Borghi e Binkofski (2014, §1.2) distinguono radicalmente la astrazione (*abstraction*) dei concetti sovraordinati dalla peculiare astrattezza (*abstractness*) dei concetti astratti caratterizzati da maggiore complessità, maggiore distacco dall'esperienza fisica, maggiore variabilità rispetto agli altri tipi di concetti. I primi "appartengono a domini differenti e non hanno referenti singoli ben delimitati ed esperibili in modo diretto" (Caruana, Borghi 2016, 174). Concetti come ANIMALI e MOBILI sono più astratti di *cane* e *sedia*, ma i loro membri sono sempre istanze concrete perché ricadono sul livello di base. I concetti astratti invece non sono percepibili ai nostri sensi.

L'astrazione è un processo mediante il quale la "conoscenza di una specifica categoria" forma una rappresentazione globale di quest'ultima (Barsalou 2003, 389). Riguarda tutti i livelli gerarchici, e perfino i concetti subordinati come *cocker* astraggono da singole istanze e da specifiche esperienze e non coincidono con il *token* in carne e ossa del nostro cane Fufi. La ricerca di un *grounding* dei concetti astratti nel sistema sensorimotorio è legittima, ma deve tener conto della loro *abstractness* peculiare. Noi possiamo quantificare bene gli esemplari della categoria

ANIMALE, ma non possiamo farlo con l'estrema variabilità degli eventi e stati evocati da un concetto astratto.

La distinzione è importante, dal momento che molti pensano che lo sforzo di ricondurre agli schemi di percezione e azione i processi di alto livello come il linguaggio e il pensiero sia accettabile per le parole concrete, ma non per i concetti astratti.

Per i teorici dell'*embodiment* quella del livello di base rimane pur sempre una rappresentazione *disembodied*, mentre per noi apre il concetto all'esperienza della nostra competenza referenziale lessicale (Marconi 1999) pensandolo attraverso i suoi esemplari. Velardi (2005) ha ipotizzato una teoria del sovraordinato vuoto e pieno, a seconda che esso sia pensato più astrattamente attraverso una lista proposizionale di caratteristiche (*features*) o, tangibilmente, attraverso gli esemplari del *livello di base***,** profilando un duplice approccio analitico e non analitico, intensionale o estensionale, ai concetti (Brooks 1987, 2005; Norman *et al.* 2007). Anche i dati sui deficit neuropsicologici specifici per categoria relativi alla afasia e alla demenza semantica (vedi sotto §4) confortano questa teoria che, nella nostra visione, contiene sia la prospettiva multimodale che quella amodale. Le teorie *embodied* invece "suggeriscono come i concetti sovraordinati sono rappresentati da una collezione di esemplari e un contesto abituale aiuta la comprensione delle loro caratteristiche comuni", mentre "il contesto non è così necessario per permettere di riconoscere le singole istanze delle categorie del livello di base come cani e sedie" (Borghi, Binkofski 2014, 5).

In Murphy e Wisniewski (1989) il ruolo del contesto è più forte per il riconoscimento dei concetti sovraordinati (ANIMALI, MOBILI, STRUMENTI MUSICALI) che non per gli oggetti di base corrispondenti (*leone*, *divano*, *chitarra*). In Borghi *et al.* (2005) le parole di sovraordinati

(FRUTTA) danno risposte più veloci con *locations* più ampie (*scenes*), che possono contenere più esemplari, come un paesaggio di campagna, in relazione alla coppia FRUTTA/arancia, mentre i concetti di base (*arancia*), danno risposte più veloci con *locations* più piccole e adatte agli oggetti come un cesto. Il contesto dunque è una sorta di collante che unisce insieme i differenti esemplari di una categoria. Kalénine *et al.* (2009) conferma l'ipotesi con un compito di categorizzazione con termini basici o sovraordinati, del tipo ''*a kind of*?'', somministrato a bambini, tra 7 e 9 anni, e adulti. L'immagine *target* (per esempio una *ciotola* accompagnata dalla domanda *A kind of bowl/utensil*?) è preceduta da un *action prime* (una fotografia di una mano nell'atto di afferrare) o da un *context prime* (la fotografia di una scena, per esempio la sala da pranzo). Il vantaggio del livello di base sul livello sovraordinato è più grande nella condizione dell'*action prime* che del *context prime*.

Questi dati mostrano l'estrema sensibilità del sovraordinato alla dimensione concreta e contestuale e suggeriscono che bisogna andare oltre un'idea troppo amodale di astrazione, trovando una compatibilità fra rappresentazione e *embodiment.*

Il concetto è multimodale e implica informazioni di tipo visivo, tattile, uditivo, motorio e finanche affettivo–emotive sulle nostre interazioni con gli esemplari di questa specie. I concetti non sono simboli astratti, ma *simboli percettivi*. Barsalou (1999) ha aggiunto alla multimodalità l'idea della simulazione per segnalare che la generazione dei concetti non trasferisce l'esperienza in un formato differente, ma innesca la riattivazione parziale di esperienze precedenti. Altre teorie sottolineano gli aspetti motori della simulazione fondandosi sull'evidenza che la lettura di parole come *calciare* innesca nella mente la programma-

zione dello schema motorio attivando piede e gamba. Si pensi alla radicale teoria dell'*Action Based Learning* (ABL) secondo cui l'ascolto di verbi o frasi di azione induce nell'ascoltatore l'attivazione di programmi motori corrispondenti, riconducendo la spiegazione dei significati ad una *perfomance* pratica più che ad descrizione semantica teorica (Glenberg, Gallese 2012).

2. Rappresentazione, *embodiment* e concetti astratti

I concetti astratti non pongono un problema univoco e coeso, ma una serie di problemi diversificati e articolati (Barsalou 2010) e forse solo una molteplicità di approcci e una teoria della rappresentazione multipla li spiegano adeguatamente (Dove 2016).

Uno dei problemi principali da spiegare è l'effetto concretezza per cui c'è un vantaggio delle parole concrete su quelle astratte rispetto al tempo di elaborazione e al ricordo. Per la teoria della *disponibilità del contesto* le parole concrete sono associate in modo debole a parecchi contesti; per la teoria del *doppio codice* esse sono rappresentate da un doppio sistema di immagini e di informazioni linguistiche (Paivio 1987). Recenti ricerche neurofisiologiche hanno utilizzato gli effetti di concretezza mostrando come le parole concrete hanno attivazione bilaterale e le parole astratte attivano solo l'emisfero sinistro. Queste ultime sono quindi maggiormente legate a informazioni di tipo linguistico.

La teoria del doppio codice fallirebbe anche perché, al contrario della sua predizione, anche i concetti astratti attivano informazioni sensorimotorie come mostrato dall'*Action–Sentence Compatibility Effect* (Glenberg *et al*. 2008), per cui frasi contenenti verbi di trasferimento come

dare e *prendere,* applicati a parole concrete (dare/prendere una pizza) o a parole astratte (dare/ricevere una notizia), elicitano risposte più veloci se la direzione del movimento del tasto di risposta è compatibile con la direzione del verbo, in coerenza con la tesi per cui comprensione del linguaggio e pianificazione del movimento siano a carico di un unico sistema sensorimotorio. Cosa che vale sia per i concetti concreti che per quelli astratti, mostrando una piena continuità nella tipologia dei concetti.

Uno dei limiti di queste ricerche è che trovare correlazioni statistiche simili per i due tipi di concetti, non vuol dire renderli cognitivamente equiparabili. Come vedremo inoltre, permangono delle differenze per le quali i concetti astratti attivano altri sistemi in aggiunta al sistema sensorimotorio in coerenza con i dati per cui parole concrete e parole astratte attivano network neurali differenti e i secondi attivano maggiormente l'emisfero sinistro. Le parole astratte inoltre tendono a richiamare relazioni associative come armadio–vestito, furto–punizione, mentre le parole concrete relazioni di similarità come sedia–poltrona, furto–scasso. Alcune sindromi o disturbi cognitivi e neuropsicologici fanno emergere difficoltà specifiche per le parole astratte. La dislessia profonda induce errori di lettura ad alta voce più per le parole astratte che per le parole concrete. Differenti *performance* per i tipi di concetti vengono esibite nella demenza semantica.

Per questo motivo alcuni teorici *embodied* hanno cercato di mantenere una certa differenza. La teoria *situazionale e introspettiva* di (Barsalou, Wiemer–Hastings, 2005), definisce i concetti astratti non solo come mancanti di referenti singoli percepibili, ma in positivo come quei concetti che contengono informazioni di tipo introspettivo ed emotivo con un sistema *embodied* molto diverso dai concetti concreti. Il paradigma di generazione delle proprietà dei

concetti rileva che i concetti concreti come *frutta* elicitano situazioni, i concetti astratti come *verità* elicitano aspetti istituzionali e sociali di queste situazioni e informazioni introspettive. Per la cui la verità elicita finanche "difficile da discutere dopo il Postmodernismo"! Vi è però il caso dei numeri, che elicitano *affordance*, mostrando che la teoria non è generalizzabile a tutti i concetti.

L'*Affective Embodied Account* di Vigliocco *et al.* (2014) mostra il collegamento con le emozioni e l'attivazione del cingolo anteriore, pur rimanendo il problema che molte delle sue evidenze potrebbero dipendere dalla presenza dei concetti emotivi all'interno degli *items* dei concetti astratti proposti negli esperimenti.

Un candidato molto incisivo è la teoria della *metafora concettuale* secondo cui il corpo modella il nostro processing dei concetti astratti attraverso un *mapping* che proietta le informazioni fisico–motorie di un dominio source, che fungono da funzioni di un modello, sul dominio *target* di un concetto astratto che viene rivisualizzato e gestito totalmente attraverso il modello concreto (Lakoff, Johnson 1980). Questo *mapping* non è in grado di esaurire tutte le caratteristiche di un concetto astratto. È però un'evidenza che, attraverso metafore centrate sul corpo, noi possiamo elaborare concettualizzazioni come L'AMORE È UN VIAGGIO, LA VITA VA ABITATA (dove è in atto una spazializzazione del tempo), LE ORGANIZZAZIONI SOCIALI SONO PIANTE, L'AMORE È UNA GUERRA. La proiezione aiuta a comprendere le frasi in cui utilizziamo questa metafora concettuale. Si instaurano collegamenti tra domini cognitivi astratti e domini visuo–motori quali *tempo*/*movimento*, *cause*/*forze fisiche*, *categorie*/*contenitori*. Le radici di questo legame stanno nel fatto che nei bambini il dominio visivo è indistinto dal dominio cognitivo astratto, la prima conoscenza è tutta di

tipo tattile e visivo permettendo in seguito molte estensioni metaforiche. Già dalla nascita inoltre essi associano esperienze sensorimotorie con esperienze relazionali–emotive come nell'abbraccio materno per cui la sensazione di calore viene in seguito associata ai *calorosi* applausi, al *caloroso* benvenuto etc. Va ricordato però che nello sviluppo l'uso delle metafore arriva più tardi e segue quello dell'utilizzo dei concetti astratti. Per questo si dovrebbe supporre un *processing* infantile più emancipato dalla modalità specifica e dunque più amodale (Caruana, Borghi 2016).

In generale il vocabolario sensorimotorio aiuta i parlanti nella comprensione delle parole astratte per cui noi diciamo "intravedere una soluzione", "essere colpiti da un modo di fare", "afferrare idee". La struttura del dominio sorgente viene trasmessa al dominio *target* e quindi gli argomenti previsti dal verbo concreto, per esempio *afferrare*, strutturano il dominio delle idee che devono quindi "essere possedute da un agente" il quale può anche decidere di "allentare la presa" o di "disfarsene". Notevole è il caso della spazializzazione del tempo (Casasanto, Boroditsky 2008), anche se si è scoperto che le metafore del tempo sono influenzate dalle culture per cui gli italiani preferiscono la prospettiva *ego–moving*, mentre i cinesi quella *time–moving* e che la valutazione del flusso è influenzata dalla quantità per i greci (un incontro *largo*), mentre è influenzata dalla durata per gli inglesi (un incontro *lungo*) (Casasanto 2008) e che *gli occidentali prediligono la metafora avanti–dietro, mentre gli orientali quella sopra–sotto per cui il futuro non è avanti, ma sopra!*

Dove (2016) segnala altri tre limiti di questa teoria: il problema della generalizzazione emerge in contesti non metaforici; non tutte le metafore linguistiche implicano sistemi percettivi; non tutti i concetti astratti permettono un

mapping col dominio concreto. Si pensi a parole come fisica, chimica. Anche se lo immaginassimo per i concetti di *verità* e *democrazia* non potremmo esaurire così la loro complessità teorica. La nozione di similarità per esempio non può essere spiegata solo attraverso il tratto della contiguità spaziale.

3. Teorie ibride della rappresentazione multipla

I limiti cui abbiamo accennato hanno portato alla costruzione di teorie delle realizzazioni multiple che non oppongono le teorie *embodied* alle teorie distribuzionali associative del significato e non assolutizzano la base dell'informazione sensorimotoria ed emozionale, rivalutando così il contributo fornito dal linguaggio. Le teorie si differenziano proprio per il ruolo attribuito a quest'ultimo. La LASS (*Language and Situated Simulation*) prevede due stadi di elaborazione (Barsalou *et al.* 2008). Quando ci viene detta la parola *cane* noi attiviamo innanzitutto una rete di parole associate *gatto*, *osso*, *cuccia*, *padrone*, *guinzaglio*. Poi iniziamo a pensare a degli specifici cani attivando esperienze avute con loro. Il linguaggio permette l'accesso alla simulazione e solo quando questa emerge noi comprendiamo davvero il significato di una parola. I verbi *pensare*, *dubitare*, *persuadere* attivano aree legate al mentale, mentre il concetto di chimica attiva aree relative agli elementi e alle reazioni. I compiti di decisione lessicale attivano informazione linguistica, mentre compiti immaginativi attivano informazione sensorimotoria. Le evidenze della LASS sono molto scarse e l'idea di una precedenza della informazione linguistica sul sensorimotorio contrasta con i dati di compiti di decisione lessicale in cui verbi di azione eseguite con arti inferiori, superiore e con

la bocca (*calciare*, *sollevare*, *leccare*) attivano le aree motorie nel giro di 150 *ms* con una corrispondenza precisa con le regioni somatotopiche dell'omuncolo (Pulvermüller *et al.*, 2005). Da qui revisioni della teoria secondo cui possiamo assistere ad oscillazioni nella successione delle informazioni. Rimane lo sforzo però di vedere il complesso contributo alla concettualizzazione.

Il filosofo Guy Dove (2009, 2011, 2014) si è appellato al principio del "pluralismo rappresentazionale" nella sua teoria delle realizzazioni multiple. Il *processing* dei concetti astratti risiede su molteplici sistemi di rappresentazione di tipo sensorimotorio e linguistico. Se i concetti concreti si basano su immagini mentali e informazioni *embodied*, i concetti astratti richiedono invece uno strumento più potente in grado di fornire una generalizzazione e questa non può che tendere all'amodalità.

Dove (2016) ricorre a studi di *neuroimaging* per i quali i concetti astratti attivano aree più lateralizzate a sinistra e collocano il *processing* concettuale in un *hub* amodale collocato nei lobi temporali anteriori bilaterali, dove si presenta la degenerazione che porta alla demenza semantica (Jefferies, Lambon Ralph, 2006, Lambon Ralph *et al.*, 2010; Pobric *et al*., 2010). Il riferimento è alla teoria *hub–and–spoke* (Lambon Ralph *et al.*, 2009; Patterson *et al.* 2007), che prevede centri (*hub*) amodali con connessioni con terminali periferici (*spoke*) sensoriali a modalità specifica. Esisterebbe dunque un'area dei concetti astratti indipendente da una specifica modalità sensoriale o motoria. Alcune teorie identificano diversi *hubs* con una visione dinamica delle interazioni con gli *spoke*. La demenza semantica provoca un deficit specifico per i concetti astratti colpendo l'*hub* amodale. In realtà vengono preservati concetti numerici mostrando delle differenze nel dominio dei concetti astratti come già segnalato sopra (§2) citando co-

me controesempio della teoria introspettiva il caso dei numeri che elicitano *affordance* e non stati interni. La localizzazione è controversa perché solo le evidenze con la TMS (*Transcranial Magnetic Stimulation*) o con esperimenti comportamentali sarebbero utili per una disconferma dell'*embodiment*. Si discute anche la nozione di "area associativa amodale" come "area celebrale adibita a una codifica semantica astratta e svincolata dai vari formati sensoriali, sostenendo viceversa che le cosiddette aree associative siano piuttosto regioni cerebrali in cui confluiscono e vengono integrate informazioni da molteplici aree sensoriali e siano dunque regioni non *a*–modali, bensì *multi*–modali" (Caruana, Borghi 2016, 189). Viene riconosciuto che il limite delle teorie delle rappresentazioni multiple non è quello di postulare sistemi differenti, ma rappresentazioni amodali. La teoria di Dove è dunque ibrida, ma non interamente *embodied*. Occorre invece una teoria basata su strategie multiple di *processing* tutte *embodied*. Si avanza così l'idea che il linguaggio stesso non usi simboli amodali ma aperti all'esperienza sociale multimodale.

Un tentativo interessante è quello di Jesse Prinz (2005; 2012), ideatore con Barsalou della teoria dei simboli percettivi. Egli rivaluta le suggestioni dell'empirismo di Locke, Berkeley e Hume, da noi discusse in §1, suggerendo provocatoriamente che la spiegazione dei concetti astratti è più complessa per le teorie distribuzionali che per quelle *embodied*. Nel suo modello i concetti astratti evocano situazioni, scenari che li incarnano in modo esemplare. *Libertà* evoca una situazione in cui abbiamo desiderato andare ad una festa o ad una gita, ma ci è stato proibito dai nostri genitori, oppure quella in cui si discute su quale partito votare in famiglia, in università o in ufficio cercando di evitare il condizionamento dei pari o di superiori in ruo-

lo (genitori, professori, capi). Il semplice atto di dividere una torta in parti eque o uguali può far sorgere una situazione legata alla *democrazia*. Queste situazioni sensori–motorie sono collegate ad emozioni e anche a pratiche linguistiche come ad esempio il confrontarsi, il domandare un parere o il fare una obiezione nella situazione della discussione sul voto. In questo modo il concetto astratto viene elaborato attraverso il concorso dell'informazione sensori–motoria, emotiva e linguistica, intendendo per linguaggio un sistema complesso interazionale e sociale fortemente incorporato e multimodale.

La teoria WAT (*Word As social Tool*) persegue questa strada e considera le parole non solo in relazioni ai referenti, ma come strumenti di azioni e sociali. L'uso di parole riferite a oggetti distanti contribuisce a mutare la nostra percezione dello spazio peripersonale. La teoria si occupa di quattro aspetti: acquisizione dei concetti astratti, loro rappresentazione nel cervello, effetti sul comportamento motorio e variabilità nelle lingue. Acquisire un concetto astratto si basa sull'applicazione del nome al referente come nei casi di *bottiglia*, *bambola*, *chiave* per cui è dimostrato che basta anche l'ostensione, semplice o ripetuta, del referente con la pronuncia del nome da parte del genitore. Per acquisire la parola *libertà* e *democrazia* non ci sono referenti e non bastano nemmeno esempi di situazioni paradigmatiche, ma occorre che qualcuno ci spieghi linguisticamente il senso del termine. Mentre all'inizio la categorizzazione dei bambini si fonda sulla salienza percettiva, piano piano gli stimoli sociali assumono un ruolo più preponderante. La comprensione delle parole astratte emerge così attorno ai 10–14 mesi, assieme al sorgere di capacità sociali quali il *gaze following* e la *joint attention*. I bambini di 3–4 anni sono molto curiosi e fanno molte domande sui concetti astratti. I genitori pronunciano paro-

le astratte ai bambini in età pre–linguistica indipendentemente dalla presenza di un referente. Le prime parole concrete sono apprese grazie all'ostensione. Poi parole più sofisticate come *tundra* e *scalpello* vengono acquisite grazie a sistemi multipli, percettivi o linguistici, a seconda del contesto.

Inteso in questo senso molto *grounded* ed esperienziale, il linguaggio è alla base dell'acquisizione dei concetti astratti permettendo una gestione migliore della complessa *abstractness* di questi concetti i cui membri e scenari sono molto più differenziati e variabili degli oggetti di base dei sovraordinati.

Sia i concetti concreti che quelli astratti attivano le aree sensorimotorie, ma i primi attivano maggiormente le aree linguistiche e sociali, rispondendo alla critica per cui il loro *processing* è nell'emisfero sinistro, attivando un circuito fronto–temporale dedicato all'elaborazione sintattica e semantica, con aree della corteccia temporale sinistra e del giro inferiore frontale sinistro.

L'importanza e relatività del linguaggio è confermata dai limiti nella variabilità linguistica del *processing* dei concetti concreti sottolineata dall'ormai famoso studio di Malt *et al.* (1999) sul concetto di "contenitore". Ai soggetti di lingua spagnola, cinese, inglese vengono proposte immagini di contenitore che differivano per estensione a seconda che fossero presentate con la parola di una delle tre lingue. Nel compito di raggruppamento delle immagini le differenze cross–culturali sparivano. Abbiamo visto sopra invece come esse siano molto sensibili per la concettualizzazione della temporalità. La WAT ipotizza che, quando lo spazio degli stimoli è più strutturato, allora l'influenza del linguaggio è meno marcata.

4. Problema della costitutività, rilancio della rappresentazione e *embodiment* debole

La teoria WAT cerca di rispondere alle critiche di Dove e altri recuperando il ruolo dell'informazione linguistica senza dover ricorrere ad un'area e ad una rappresentazione amodale, ma reinterpretando i centri astratti della teoria *hub–and–spoke* in senso multimodale. Permangono però dei problemi che spingono a considerare ancora spinosa l'opposizione tra amodale e multimodale e richiederebbero l'amodalità come spiegazione del *processing* concettuale. Possiamo segnalare quattro ordini di problemi, tre interni alla teoria dei concetti e uno riguardante tutta la teoria dell'*embodiment* in relazione alla nozione di rappresentazione.

Dove (2016) segnala i tre problemi decisivi della generalizzazione astrattiva, del *disembodiment* diversificato e maggiore non solo tra i concetti concreti e i concetti astratti, ma anche tra i concetti astratti legati alla *abstracteness* (come *libertà* e *democrazia*) e i concetti astratti legati alla *abstraction* (come i concetti sovraordinati ANIMALE e FRUTTA); dell'estrema *flessibilità* dei concetti astratti. A questi si aggiunge il fatale problema della *costitutività causale* della dimensione sensorimotoria per il *processing* linguistico e concettuale (Mahon 2015a, 2015b; Mahon, Caramazza 2008). La loro discussione e una serie di evidenze robuste farebbero propendere per una integrazione dell'amodalità con l'*embodiment* e per una elaborazione a due stadi in cui ad una prima fase di costruzione concettuale multimodale segua una emancipazione più amodale della rappresentazione.

4.1 Generalizzazione e disembodiment

La generalizzazione concettuale riguarda sia la dimensione orizzontale (struttura prototipica della categoria), che la dimensione verticale dei livelli di astrazione (§1). Esse sono intrecciate e forniscono le informazioni sulla realtà andando oltre l'esperienza immediata. Per Dove (2016) questa generalizzazione è il problema universale dei concetti.

Discutendo in §1 la distinzione tra *abstractness* e *abstraction* (Borghi, Binkofski 2014) abbiamo visto che i concetti astratti non possono essere definiti a partire dalla dimensione verticale. Dove (2016) vede intrecciate *abstraction* e *abstractness* anche per i concetti astratti[6]. Il primo riguarderebbe la *generalizzazione*, il secondo il *disembodiment*. Le teorie *grounded* non riuscirebbero ad affrontare la *generalizzazione*. Patterson *et al.* (2007, 977) pensano che se la memoria semantica consiste solo di contenuti relativi ad oggetti specifici per modalità, allora non si riesce a capire come si possa pervenire a generalizzazioni di alto livello. Le rappresentazioni amodali offrono mezzi migliori per integrare informazioni a partire da molteplici fonti (Dove, 2009; Machery, 2007). Meteyard, *et al.* (2012) mostrano l'*embodied cognition* come un *continuum* differenziato da maggiore a minore intensità di *embodiment* da teorie che limitano fortemente la cognizione ai sistemi dell'azione, dell'emozione, della percezione a teorie parzialmente *grounded* nei sistemi dell'azione e della percezione fino a teorie più deboli che vedono l'attivazione sensorimotoria come una conseguenza secondaria del *processing* cognitivo diretto da aree amodali.

Dove (2016) indica due gruppi di evidenze che implicherebbero rappresentazioni amodali nella memoria semantica. Il primo viene da pazienti neuropsicologici con disturbi come la demenza semantica (SD) caratterizzata da

[6] Lo riconoscono anche CARUANA e BORGHI (2016, 174).

un'atrofia graduale bilaterale dei lobi temporali e da un concomitante e progressivo deperimento della memoria semantica per gli oggetti comuni (Lambon Ralph *et al.*, 2010). Questi soggetti presentano una conoscenza degradata di diversi *items* di vaste categoria, preservandone la conoscenza di altri, come la paziente studiata longitudinalmente con problemi di denominazione per le immagini di *acquila* e *ostrica*, ma non di *pollo*, *anitra* e *cigno* (Patterson *et al.*, 2007). Il deficit è tipicamente cross–modale e la degradazione semantica procede spesso in una maniera gerarchica dal maggiore al minore livello di astrazione. La paziente ha perso gradualmente l'abilità di identificare specie di uccelli, ma rimane capace di identificare la gran parte di loro come *uccelli* e come *mammiferi* con un utilizzo strategico del sovraordinato. Questo potrebbe suggerire che sia un *hub* amodale ad avere un ruolo centrale nella memoria semantica (McCaffrey, 2015; McCaffrey, Machery, 2012; Patterson *et al*., 2007; Reilly *et al*., 2014). Nella demenza semantica emerge un deficit per le parole concrete/*high–imageable* (e si discute se questo sia un tratto tipico della demenza semantica o no.

Il secondo gruppo di evidenze implica soggetti neurotipici. Binder *et al*. (2009) forniscono una metanalisi di 120 studi di *neuroimaging* sulle differenze tra compiti semantici e non semantici trovando numerose aree eteromodali temporali e parietali sinistre impiegate regolarmente in questi compiti. Una pletora di aree non specifiche per modalità sono implicate nel *processing* semantico. Questo corpo di evidenze rivendica un approccio amodale tradizionale o una teoria ibrida come la teoria *hub–and–spoke* che attribuisce *processing* amodale ai lobi temporali anteriori. Per Dove (2016) quest'ultima risolverebbe il problema della generalizzazione, della flessibilità e del *disembodiment*. La flessibilità sarebbe il risultato di diffe-

renti impieghi degli *spokes* a seconda del compito e del contesto. Il *disembodiment* verrebbe spiegato attraverso la differente intensità di elaborazione e di influenza degli *hub*.

4.2 *Problema della costitutività e embodiment debole*

Come abbiamo più volte ricordato alcuni teorici hanno criticato le posizioni dell'*embodiment* radicale, soprattutto in relazione alla necessarietà dell'attivazione del sistema motorio all'interno del *processing* cognitivo (Mahon 2015a, 2015b; Mahon, Caramazza 2008). Si è avanzato il cosiddetto "problema della costitutività". L'influenza dell'informazione sensorimotoria *embodied* nei processi cognitivi non sarebbe tale da implicare un suo ruolo causale centralmente e radicalmente costitutivo del *processing* concettuale che conserverebbe la sua autonomia. Il reclutamento del sistema sensorimotorio nei compiti cognitivi non è una condizione necessaria al compito cognitivo stesso, ma solo qualcosa di concomitante che interagisce con l'elaborazione mentale dei concetti.

A nostro modo di vedere l'argomento della costitutività può essere usato per sconfessare l'*embodiment* radicale, ma non per tornare a posizioni di scienza cognitiva classica di stampo fodoriano come hanno fatto Brad Mahon e Alfonso Caramazza, proponendo un *disembodied approach* (Mahon 2015a, 2015b; Mahon, Caramazza 2008). Secondo questi autori le evidenze per cui l'elaborazione del linguaggio attiva il sistema motorio non possono essere contestate. Nessuna di queste evidenze però prova che l'attivazione del sistema motorio è costitutiva dell'elaborazione delle parole. L'innesco del sistema motorio è successivo, è un'attivazione che avviene a posteriori rispetto all'elaborazione del significato linguistico che

lo precede. Prima c'è l'attivazione del linguaggio e poi quella del sistema motorio. In un modo contingente e non costitutivo. Prima comprendiamo la frase "apri la finestra" e poi successivamente decidiamo se attivare uno specifico programma motorio, a seconda del contesto in cui è avvenuta la comprensione della frase. Nel modello di Mahon e Caramazza il nucleo centrale della rappresentazione linguistica rimane astratto e amodale. L'attivazione delle cortecce sensoriali e motorie è un corollario dovuto al diffondersi del *processing* in aree non centrali per la comprensione del significato. Vi sarebbero dunque due tipi di aree: le aree attive che processano l'informazione semantica, che sono centrali ai fini della comprensione; le aree attive di secondo piano, che svolgono soltanto una pura funzione concomitante di controllo del processo centrale.

Il problema della costitutività è stringente. Ci sono evidenze che pongono l'informazione sensorimotoria in una fase immediata, come abbiamo visto criticando la teoria LASS con gli esperimenti di Pulvermüller *et al*. (2005), che pongono a 150 ms l'emergere dell'informazione motoria e altre che la ritardano a 500 ms (Papeo *et al*. 2009). I teorici dell'*embodiment* hanno cercato di *bypassare* il problema della costitutività evocando il maggiore ruolo assunto dalle aree premotorie, legate alla pianificazione, e non da quelle motorie primarie. Non a caso la denominazione di animali attiva le aree occipitali, mentre quella di utensili le aree premotorie. Un ulteriore studio di Papeo *et al*. (2014) ha studiato il giro temporale medio sinistro posteriore, regione implicata nel *processing* concettuale e distinta dalle aree sensorimotorie, fornendo evidenze che la stimolazione di questa area genera deficit sintattici per verbi di azione. Gli autori pensano che il *processing* del significato sia organizzato gerarchicamente, con rappre-

sentazioni amodali che rimangono collegate però alle aree motorie, potendo mediarne l'attivazione.

Certamente la mancata attivazione delle cortecce primarie non può portare a considerare l'informazione *embodied* come epifenomeno posticcio. La simulazione inoltre è una riattivazione parziale dell'esperienza e non un ricompierla pienamente.

Si deve quindi trovare una strada per mettere d'accordo evidenze molto diverse. E distinguere tra *embodiment* forte o debole. Nel primo caso le cortecce sensoriale e motorie hanno un ruolo causale diretto nell'elaborazione del linguaggio; nel secondo caso la rappresentazione concettuale è adiacente o anteriore all'attivazione delle cortecce sensoriali e motorie che elaborano le nostre esperienze reali.

L'insieme di problemi e di evidenze di cui abbiamo parlato induce a oltrepassare l'antitesi tra amodalità e multimodalità, tra rappresentazione ed *embodiment* e proporre un modello a due stadi di elaborazione. La prospettiva dell'*embodiment* debole potrebbe non sconfessare il ruolo con–causale dell'informazione sensorimotoria chiudendosi nell'amodalità rigida; e l'*embodiment* forte potrebbe riconoscere lo spazio adeguato per un *hub* amodale che non sia per forza multimodale in modo pervasivo.

Si potrebbe ipotizzare un processo a due stadi. Nel primo il sensorimotorio si integra con la rappresentazione che rimane legata all'esperienza con gli oggetti, riferendosi così a *token* particolari, al simbolo percettivo di Barsalou, con una concettualizzazione esperienziale già incamminata verso l'astrazione. Nel secondo la rappresentazione diventa amodale sfruttando gli *hub* presenti nel cervello di cui abbiamo parlato sopra e risolvendo il problema della *generalizzazione*. Questo secondo livello permette due scenari cognitivi simili a quelli visti per il concetto so-

vraordinato, pieno e vuoto, in §1. La rappresentazione ha un suo dominio amodale e un livello di autonomia ed emancipazione dal sensorimotorio nel quale il concetto può essere elaborato astrattamente attraverso reti distribuzionali associative legate ai legami con altre parole–concetto o con le liste di caratteristiche, in formato proposizionale, che definiscono la categoria. Questa autonomia non impedisce un continuo ritorno al dominio dell'esperienza e del sensorimotorio, i cui portati erano già rientrati nella rappresentazione amodale, che è nata come simbolo percettivo, rimanendo così disponibili all'elaborazione qualora li richieda in generale e per specifici compiti.

5. Conclusioni

La ricognizione dello stato dell'arte della teoria *embodied* dei concetti astratti e dei problemi ad essa connessi fa comprendere come né le argomentazioni e le evidenze addotte dai fautori di una teoria sensori–motoria e multimodale dei concetti, né quelle addotte dai fautori della teoria rappresentazionale amodale e astratta, possono essere rifiutate in blocco (§1, §2, §4.1). Occorre trovare una teoria intermedia e ibrida (§3, §4.2) che mostri come possano convivere la multimodalità e la amodalità all'interno di un *processing* complesso che esige un doppio stadio e due livelli autonomi di elaborazione. Il doppio stadio indica che i concetti si formano a partire da un *processing* più legato alla dimensione sensori–motoria e *bodily grounded*, in cui l'interazione corporea con gli oggetti gioca un ruolo fondamentale (§4.2).

Questa dimensione non esclude che la loro generazione maturi fino a permettere la formazione di rappresentazioni

più sganciate dalla interazione corporea e dagli schemi sensori–motori, caratterizzate dunque da prospettiva amodale che permette alla rappresentazione stessa di raggiungere un livello di piena autonomia dalla dimensione *embodied*, in cui essa può essere utilizzata dalla nostra mente attraverso una manipolazione più astratta.

In questo modo, da una parte non si nega che vi sia una dimensione sensori–motoria della concettualizzazione, dall'altra non si nega nemmeno che ci sia una dimensione amodale e astratta della rappresentazione. Non si nega nemmeno che queste due dimensioni possano essere intrecciate e interagire fra loro sia nell'ottica di una interazione tra il sensori–motorio e l'amodale, sia di una presenza e traccia della dimensione sensori–motoria nella stessa rappresentazione amodale, così come previsto per l'analogato visivo nella teoria dei prototipi. Infatti il livello in cui matura la rappresentazione amodale non si sgancia totalmente dalla dimensione sensori–motoria con cui essa era legata nel primo stadio della concettualizzazione. Ma ne conserva le tracce permettendo la permanenza di due prospettive bifronti della rappresentazione, così come accade nella teoria del sovraordinato pieno e vuoto cui abbiamo accennato in §1. In una prospettiva il concetto viene visualizzato attraverso il contributo degli schemi più *embodied* e di tipo sensori–motorio, in un'altra prospettiva la rappresentazione elabora i contenuti più astratti legati all'astrazione dei singoli esemplari in quanto occorrenze concrete (*token*), delle categorie dei vari livelli di astrazione (subordinato, di base, sovraordianto), delle proprietà e dei *features* delle categorie viste da un punto di vista più astratto–intensionale.

Come si vede questa teoria permette una maggiore correlazione tra asse estensionale (più legato agli esemplari appartenenti ad una categoria) e asse intensionale (più le-

gato alla lista delle proprietà e delle caratteristiche che definiscono gli esemplari appartenenti ad una categoria o la categoria medesima).

Occorrerà inoltre approfondire ulteriormente questa ipotesi di maggiore comunicazione tra dimensione sensori–motoria e dimensione astratta–amodale della concettualizzazione e tra dimensione degli oggetti concreti (*tokens*) e dimensione dei concetti (*types*) per la quale aggiungiamo l'osservazione che essa è rintracciabile, con varie oscillazioni, in molte teorie filosofiche della generazione delle idee e della conoscenza (Velardi 2013).

Riferimenti bibliografici

BARSALOU L. W., (1999) *Perceptual symbol systems*, «*Behavioral and Brain Sciences», 22,* 577–609.

BARSALOU L. W., (2003) *Abstraction in perceptual symbol systems*, «Philosophical Transactions of the Royal Society of London: Biological Sciences», 358, 1177–1187.

BARSALOU L. W., (2008) *Grounded cognition*, «*Annual Review of Psychology», 59,* 617–645.

BARSALOU L. W., (2010) *Grounded cognition: Past, present, and future*, «*Topics in Cognitive Science», 2,* 716–724.

BARSALOU L. W., WIEMER–HASTINGS K., (2005) *Situating abstract concepts*, in PECHER D., ZWAAN R., (Eds.*),* (2005) *Grounding cognition: The role of perception and action in memory, language, and thought*, Cambridge University Press, Cambridge, UK, pp. 129–163.

BARSALOU L. W., SANTOS A., SIMMONS W. K., WILSON C. D. (2008) *Language and simulation in conceptual processing*, in M. DE VEGA, GLENBERG A. M., GRAESSER

A.C. (eds.), (2008), *Symbols and embodiment: Debates on meaning and cognition*, Oxford University Press, New York, NY, pp. 245–284.

BORGHI A.M., BINKOFSKI, F., CASTELFRARCHI, C., CIMATTI, F., SCOROLLI, C., TUMMOLINI, L., (2017) *The challenge of abstract words*, «Psychological Bulletin», 143 (3): 263–292.

BORGHI A.M., BINKOFSKI F., (2014) *Words as social tools: an embodied view on abstract concepts*, Springer, New York/ Berlin.

BROOKS L. R. (1987) *Decentralized control of categorization: the role of prior processing episodes*, in NEISSER U., (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge University Press, Cambridge, UK, 141–174.

BROOKS *L. R.,* (2005*) The blossoms and the weeds,* «Canadian Journal of Experimental Psychology», 59(1), 62–74.

CARUANA F., BORGHI A., (2016) *Il cervello in azione*, Il Mulino, Bologna.

BINDER J. R., DESAI R. H., GRAVES W. W., CONANT, L. L. (2009) *Where is the semantic system? A critical review and meta–analysis of 120 functional neuroimaging studies*, «*Cerebral Cortex», 19,* 2767–2796.

CASASANTO D. (2008) *Who's afraid of the big bad Whorf? crosslinguistic differences in temporal language and thought*. «Language Learning», 58, 63–79.

CASASANTO D., BORODITSKY L. (2008) *Time in the mind: Using space to think about time*, «Cognition», 106, 579–593.

DOVE G., (2009) *Beyond perceptual symbols: A call for representational pluralism*, «*Cognition», 110,* 412–431.

DOVE G., (2011) *On the need for embodied and dis–embodied cognition*, «*Frontiers in Cognition», 1*(242), 1–13.

DOVE G., (2014) *Thinking in words: Language as an embodied medium of thought*, «*Topics in Cognitive Science», 6,* 371–389.

DOVE G., (2016) *Three symbol ungrounding problems: abstract concepts and the future of embodied cognition*, "Psychon. Bull. Rev.", 23, 1109–1121.

FODOR J.A., (1976*) The Language of Thought, Harvester Press, Sussex, UK.*

GALETZKA C., (2017) *The Story So Far: How Embodied Cognition Advances Our Understanding of Meaning–Making*, «Frontiers in Psychology», 8:1315.

GALLAGHER S., (2017) *Enactivist Interventions. Rethinking the Mind*, Oxford University Press, Oxford.

GALLESE V., LAKOFF G., (2005) *The brain's concepts: The role of the sensory–motor system in reason and language*, «*Cognitive Neuropsychology», 22,* 455–479.

GLENBERG A. M., GALLESE V., (2012) *Action–based language: a theory of language acquisition, comprehension, and production*. «Cortex», 48, 905.

KALÉNINE S., BONTHOUX F., BORGHI A. M., (2009) *How action and context priming influence categorization: a developmental study*, «British Journal of Developmental Psychology», 27, 717–730.

KOSSLYN S.M., (1980) *Image and Mind*, Harvard University Press, Cambridge, MA.

KOSSLYN S.M., (1983) *Ghost's in the Mind Machine*, W. W. Norton & Co., New York, NY.

KOSSLYN S.M., THOMPSON W., GANIS G., (2006) *The Case for Mental Imagery*, Oxford University Press, Oxford.

JEFFERIES E., LAMBON RALPH M. A., (2006) *Semantic impairment in stroke aphasia versus semantic dementia: A case–series comparison*, «*Brain», 129,* 2132–2147.

LAKOFF G., JOHNSON M., (1980) *Metaphors we live by*, University of Chicago Press, Chicago, IL.

LAMBON RALPH M. A., POBRIC G., JEFFERIES E., (2009) *Conceptual knowledge is underpinned by the temporal lobe bilaterally*. *Convergent evidence from rTMS*, «*Cerebral Cortex», 19,* 832–838.

LAMBON RALPH M. A., SAGE K., JONES R. W., MAYBERRY E. J., (2010) *Coherent concepts are computed in the anterior temporal lobes*, «*Proceedings of the National Academy of Sciences», 107,* 2717–2722.

MAHON B. Z., (2015a) *The burden of embodied cognition*, «Canadian Journal of Experimental Psychology», 69, 172–178.

MAHON B. Z., (2015b) *What is embodied about cognition*, «*Language, Cognition, and Neuroscience», 30,* 420–429.

MAHON B. Z., CARAMAZZA A., (2008) *A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content*, «*Journal of Physiology», 102,* 59–70.

MARCONI D., (1999) *La competenza lessicale*. *Cosa vuol dire saper usare le parole*, Laterza, Roma–Bari.

MCCAFFREY J., (2015) *Reconceiving conceptual vehicles: Lessons from semantic dementia*, «*Philosophical Psychology», 28,* 337–354.

MCCAFFREY J., MACHERY E., (2012), *Philosophical issues about concepts*, «*Wiley Interdisciplinary Reviews: Cognitive Science», 3,* 265–279.

METEYARD L., CUADRADO S. R., BAHRAMI B., VIGLIOCCO G., (2012) *Coming of age: A review of embodiment and the neuroscience of semantics*, «*Cortex», 48,* 788–804.

MURPHY G. L., *(2002) The Big Book of Concepts*, MIT Press, Cambridge MA.

MURPHY G. L., WISNIEWSKI E. J., (1989) *Categorizing objects in isolation and in scenes: What a superordinate is good for*, «Journal of Experimental Psychology. Learning, Memory, and Cognition», 15, 572–586.

NORMAN G., YOUNG M., BROOKS L., (2007) *Non–analytical models of clinical reasoning: the role of experience*, «Medical Education», 41(12):1140–5.

PAIVIO A., (1986) *Mental representations: A dual coding approach*, Oxford University Press, New York, NY.

PAPEO L., VALLESI A., ISAJA A., RUMIATI R.I., (2009) *Effects of TMS on different stages of motor and non–motor verb processing in the primary motor cortex*, PLoS ONE, 4(2): e450.

PAPEO L., LIGNAU A., AGOSTA S., PASCUAL–LEONE A., BATTELLI L., CARAMAZZA A., (2014) *The origin of word–related motor activity*, «Cerebral Cortex», 25, 1668–1675.

PATTERSON K., NESTOR P. J., ROGERS T. T., (2007) *Where do you know what you know? The representation of semantic knowledge in the human brain*, «*Nature Reviews Neuroscience», 8,* 976–987.

POBRIC G., JEFFERIES E., LAMBON RALPH M. A., (2010) *Amodal semantic representations DEPEND on both left and right anterior temporal lobes: New rTMS evidence*, «*Neuropsychologia», 48,* 1336–1342.

PRINZ J. J., (2005) *The return of concept empiricism*, in H. Cohen, C. Leferbvre (Eds.), *Categorization and cognitive science*, Elsevier, New Jersey.

PRINZ J. J., (2012) *Beyond human nature. How culture and experience shape our lives*, Penguin, London, Norton, New York.

PULVERMÜLLER F., SHTYROV Y., ILMONIEMI R., (2005) *Brain signatures of meaning access in action word recognition*, «Journal of Cognitive Neuroscience», 17, 884–892.

PYLYSHYN Z., (1973) *What the Mind's Eye Tells the Mind's Brain*, «Psychological Bulletin», 80, 1–24.

PYLYSHYN Z., (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science*, MIT Press, Cambridge MA.

PYLYSHYN Z., (2004) *Seeing and Visualizing: It's Not What You Think*, MIT Press, Cambridge MA.

REILLY J., HARNISH S., GARCIA A., HUNG J., RODRIGUEZ A. D., CROSSON B., (2014) *Lesion symptom mapping in nonfluent aphasia: Can a brain be both embodied and disembodied?*, «*Cognitive Neuropsychology», 31,* 287–312.

ROSCH E., MERVIS C.B., GRAY W., JOHNSON D., BOYES–BRAEM P., (1976) *Basic Objects in Natural Categories*, «Cognitive Psychology», 8(3), 382–439.

SEARLE J.R., (1980) *Minds*, *brains, and programs,* «Behavioral and Brain Sciences», 3(3), 417–457.

VELARDI A., (2005) *Il nuovo paradigma. Categorie, prototipi e semantica cognitiva*, EDAS, Messina.

VELARDI A., (2013) *La vita delle idee. Il problema dell'astrazione nella teoria della conoscenza*, Mimesis, Milano.

VIGLIOCCO G., KOUSTA S. T., DELLA ROSA P. A., VINSON D. P., TETTAMANTI M., DEVLIN J. T., CAPPA S. F., (2014) *The neural representation of abstract words: The role of emotion*, «*Cerebral Cortex», 24,* 1767–1777.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/978882551306613
pp. 265-298 (marzo 2018)

Concetti e produzione del linguaggio

Relazioni categoriali e iperonimia nel modello del lemma

di ANDREA VELARDI[1]

1. Introduzione

La teoria dell'accesso lessicale è una parte consistente della teoria della produzione del linguaggio e studia i processi attraverso i quali la mente elabora la verbalizzazione dei concetti che il parlante vuole esprimere traducendo il concetto astratto in una parola di una determinata lingua dotata quindi di un corredo morfo–sintattico e di una articolazione fonologica.

Sono molte le teorie che si sono susseguite nel corso di questi decenni. Uno dei problemi principali è stato quello di comprendere quanto questo processo sia seriale o parallelo. Un'altra importante questione ha riguardato il problema di come il *conceptualization problem* sia collegato al *verbalization problem* e come quindi il concetto sia tradotto nella mente in un formato che si avvicini alla parola fonologica, ma che ancora non abbia l'articolazione morfo–sintattica e sonora di quest'ultima.

Levelt *et al.* (1999) ha proposto di chiamare *lemma* questa parola astratta, già corredata da alcuni attributi

[1] Università di Messina, Dipartimento di Scienze Cognitive, Psicologiche, Pedagogiche e degli Studi Culturali COSPECS. E–mail: velardi.velardi@gmail.com

grammaticali, che si colloca nello stato intermedio tra il concetto e la parola fonologicamente articolata nella lingua dei parlanti. Il modello del *lemma* è divenuto uno principali modelli nel campo della ricerca in quanto risponde adeguatamente ai problemi sollevati dalla spiegazione di numerosi fenomeni psico–linguistici tra cui quelli del *tip–of–the–tongue* (TOT) e della denominazione di figure in pazienti normali e neuropsicologici (soprattutto afasici o affetti da demenza semantica) nei quali si assiste alla capacità di indicare il nome della categoria superiore di appartenenza dell'esemplare rappresentato nella figura in una comunicazione dei livelli di astrazione dei concetti che rende più plastico il legame tra sovraordinato, categoria di base e oggetto (si veda l'altro saggio pubblicato in questo volume).

In particolare il modello, rivisitato e ampliato da Roelofs (1992, 1997, 2014) ha focalizzato ampiamente la questione delle relazioni tra i livelli di astrazione di una parola–concetto, enunciando il cosiddetto problema dell'iperonimia *(hypernym problem*) o della ereditarietà dei tratti (Levelt 1989, 201–214, Roelofs 1992, 1997) considerato uno dei problemi più spinosi, un vero e proprio banco di prova (*touchstone*) per il passaggio dalla concettualizzazione alla verbalizzazione e dunque per tutta la teoria dell'accesso lessicale. Esso riguarda la *convergenza* del parlante verso quella peculiare parola, situata in un preciso livello di astrazione, che viene considerata più pertinente per esprimere le intenzioni comunicative dei soggetti e l'aderenza al contesto in cui viene enunciato il messaggio (motivo per cui, davanti all'animale in gabbia allo zoo, noi diciamo "Quel leone", e non semplicemente "Quell'ANIMALE", oppure indichiamo preferibilmente "Il nostro GENITORE" con "Mio padre" e non con il sovraordinato che utilizziamo per definire *padre* e *madre*

(§4,1). Questa convergenza *bypassa* la necessità di fare ereditare alla categoria di ordine inferiore i tratti e i componenti che definiscono o caratterizzano la categoria di ordine superiore così come prevede un modello logico–normativo astratto. Quando il parlante dice *leone* non intende per forza dire ANIMALE trascinando su *leone* tutte le proprietà previste da una definizione di ANIMALE che, comunque, la categoria di base deve condividere con il suo sovraordinato in una prospettiva logico–astratta.

Una teoria della produzione lessicale richiede quindi di affrontare al meglio il problema della gerarchia semantica delle categorie e dei livelli di astrazione per rendere conto del modo complesso, e altamente adattativo ai contesti e alle intenzioni soggettive, con cui i parlanti denominano gli oggetti della realtà intendendo riferirsi di volta in volta ad una concettualizzazione più o meno focalizzata degli oggetti denominati con meccanismi sofisticati di *zoom in* e *zoom out* (Velardi 2005, Velardi 2013).

Per converso una teoria dell'accesso lessicale è fondamentale per l'elaborazione di una teoria linguistico–cognitiva del significato matura e aperta alla forza delle evidenze sperimentali delle neuroscienze. Anche se occorre che pure i neuroscienziati si confrontino con la complessità dei modelli teorici proposti dalla linguistica (Perconti 2001, 180–183).

Per indagare al meglio la relazione tra *conceptualization problem* e *verbalization problem* approfondiremo innanzitutto il modello del lemma nelle sue versioni classiche e nelle successive applicazioni (§§2, 3). Ci concentreremo anche sui problemi che lo riguardano e sul dibattito suscitato in letteratura. Tratteremo poi il problema dell'iperonimia e le soluzioni fornite da Levelt e Roelofs (§4.1) Approfondiremo soprattutto la proposta di quest'ultimo che elabora una teoria che, oltre

all'applicazione del principio della *spreading activation* (Collins, Loftus 1975), da preferire ai modelli di tipo gerarchico (Collins, Quillian 1969), propone, riprendendo alcune suggestioni di Fodor (Fodor 1976, Fodor *et al.* 1975; Fodor *et al.* 1980) e criticando Jackendoff (1983, 1987, 1990), un modello non componenziale, non analitica, sintetico–olistica del significato e dei concetti (§4.2, §4.3), in cui la dimensione componenziale e analitica viene lasciata alla prospettiva più semantica delle definizioni e del *processing* meramente concettuale dei significati, vedendo come più pertinente la non componenzialità (o composizionalità) per il *processing* della produzione linguistica e del *lexical access* relativo alla verbalizzazione dei concetti stessi. Viene rifiutata la radicalità con cui Jackendoff sovrappone l'analisi componenziale del concetto alla scomposizione lessicale della parola, e viene preferito l'approccio più moderato di Fodor che prevede un approccio analitico per i concetti e uno olistico per le parole. In questo modo la teoria non composizionale di Roelofs (1997, 2014) emerge da una approfondita e articolata discussione delle teorie componenziali e non–componenziali che hanno attraversato la riflessione linguistico–concettuale del Novecento (§4.2).

Mostreremo l'importanza di questo dibattito, l'utilità di questa prospettiva olistica e i problemi ad essa connessi (§4.3) legati al fatto che il parlante mostra una grande elasticità e libertà nella espressione verbale dei concetti intrecciando o separando la prospettiva analitica e quella olistica a seconda dei contesti e delle proprie intenzioni con variabili di tipo cognitivo, linguistico–semantico e pragmatico che vanno indagate più approfonditamente in ulteriori ricerche teorico–sperimentali.

2. Accesso lessicale e teoria del lemma

Levelt *et al*. (1999) hanno proposto un modello computazionale chiamato WEAVER ++ (*Word from Encoding by Activation and VER-ification*) per la produzione di singole parole. Gli assunti del modello sono i seguenti: esiste una rete di diffusione in avanti dell'attivazione all'interno di una rete computazionale dal carattere discreto e seriale, poi modificato in senso più distribuito e parallelo, in cui si giunge al passaggio successivo solo se si è completato il passaggio precedente. Vi sono sei fasi: preparazione concettuale (concetto lessicale), selezione o accesso lessicale (lemma), codifica morfologica (forma della parola), codifica fonologica (parola fonologica), codifica fonetica (sensazione gestuale fonetica), articolazione (sistema motorio che produce le onde sonore).

I lemmi sono rappresentazioni astratte di parole che contengono proprietà sintattiche, ma non ancora proprietà fonologiche. Quando un parlante impara una lingua straniera, può accadere che conosce il significato della parola da esprimere, sa che si tratta di un sostantivo, ma non conosce la sua esatta pronuncia. In questo caso possiamo dire che ha accesso al lemma, ma non al lessema del dizionario. Un caso interessante è quello M.D., paziente con afasia globale, con un deficit categoriale specifico cross–modale nella denominazione di frutta e ortaggi (Hart *et al*. 1985). M.D. ha accesso al concetto, comprende i nomi, ma non recupera la forma fonica della parola corrispondente.

La distinzione tra lemma e forma lessicale poggia sugli studi sugli errori del parlato e sulla esistenza di errori morfologici. Come dicevamo il lemma contiene già delle indicazioni grammaticali. Il fenomeno del *tip–of–the–tongue* (TOT) è da sempre stato considerato una delle evidenze più forti per il modello. Si ha la chiara sensazione di per-

cepire il concetto e di accedere al lemma, senza che lo si riesca a tradurre in forma sonora, cercando invano la parola fonica giusta che etichetta con precisione il concetto. Vigliocco *et al.* (1997) e van Turennout *et al.* (1998) dimostrano che i soggetti, durante il TOT, sanno indicare il genere della parola che non riescono a pronunciare. Risultati meno coerenti sono ottenuti da Biedermann *et al.* (2008) in cui i soggetti non hanno accesso al primo fonema della parola dopo avere avuto accesso alla informazione di genere così come prevederebbe il modello del lemma. Abrams (2008) ha mostrato che un *priming* fonologico aiuta le persone a recuperare la parola bersaglio, ipotizzando che il TOT dipenda da una mancata rappresentazione fonologica. Starreveld, La Heij (2004) danno evidenza che il genere grammaticale di un nome è recuperato a partire dalla forma fonologica e non dal lemma.

Caramazza (1997) mette in dubbio l'esistenza del lemma. In particolare Caramazza, Miozzo (1997) si basano sul fatto che gli afasici compiono errori di sostituzione (*cane* per *gatto*) nella modalità parlata, ma non scritta, mentre in altri pazienti accade il contrario. E così alcuni pazienti hanno difficoltà a pronunciare parole e non a scriverle. Il lemma sarebbe dunque a modalità specifica e non neutra. Roelofs *et al*. (1998) preferiscono un resoconto differente, basato sul modello WEAVER ++, secondo cui il danno cerebrale influenza differentemente le connessioni tra lemma e forme parlate e scritte (Roelofs, Ferreira, *in press*, 9–10). Caramazza e Miozzo discutono inoltre il fenomeno TOT che richiede, secondo loro, solo che le informazioni grammaticali e fonologiche siano rappresentate indipendentemente, ma questo può essere fatto anche da un modello a due stadi con un solo nodo lessicale anziché due, uno per il lemma e uno per il lessema. In effetti, come abbiamo visto dai pochi esempi enunciati sopra, si può pre-

vedere che l'attribuzione delle caratteristiche grammaticali avvenga in uno stadio che preveda già la configurazione di qualcosa di molto più vicino alla parola del lemma e cioè di un lessema con informazioni grammaticali cui poi si aggiunge poi la parola articolata fonologicamente cioè il vero e proprio *output* fonatorio vocale (Perconti 2001).

Altri problemi riguardano la serialità e discretezza del modello. Una serie di evidenze mostra che l'attivazione della forma delle parole non dipende direttamente dalla selezione del lemma e sono più coerenti con il modello di diffusione dell'attivazione di Dell (1986) che prevede che l'accesso alla informazione fonologica possa iniziare prima che si completi la selezione del lemma. Vi è dunque una maggiore interazione tra i livelli (Meyer, Damian 2007; Smith, Wheeldon 2004, che riportano anche evidenze affermative) rispetto alla discretezza di Levelt *et al*. (1999).

Morsella, Miozzo (2002) dimostrano che, in un compito di denominazione della figura in verde con sovraimpressa una immagine distrattore in rosso, il tempo di denominazione è più corto quando i nomi delle immagini condividono parti della loro forma (*cat* e *cap*) rispetto a quando questo non avviene (*cat* e *pen*). Questo suggerisce che l'attivazione procede a cascata dai concetti (GATTO, TAPPO, PENNA) ai lemmi e alle forme corrispondenti, cosa che accelera la codifica della forma target (*gatto*). Per questi motivi Roelofs (2008) supera la discretezza del modello e il lemma emerge per selezione competitiva in una rete a *diffusione dell'attivazione* più parallela (Roelofs, Ferreira, *in stampa*, 9–11). *Resta che* la meta–analisi condotta su dozzine di studi da Indefrey, Levelt (2004) conferma le sei sequenze temporali del modello. La selezione lessicale avverrebbe circa 175ms dalla presentazione delle immagini e il suono appropriato verrebbe recuperato in-

torno a 250–300 ms, mentre la parola fonologica si presenterebbe a circa 455 ms e solo in seguito si attiverebbero le aree senso–motorie implicate nell'articolazione. La selezione delle parole astratte (lemmi) verrebbe completata prima. Nonostante le evidenze contrastanti il modello di Levelt ha sicuri vantaggi esplicativi e un buon grado di approssimazione con quello che accade nella realtà, senza incistarsi sull'analisi degli errori linguistici.

Roelofs, Ferreira (*in stampa*) mantengono l'assunto che la produzione si basi su tre processi principali: concettualizzazione, formulazione e articolazione (Garrett, 1975; Levelt, 1989). *Conceptualization problem* e *verbalization problem* sono estremamente connessi fra di loro. La frase *il gatto è sotto la tavola* che esprime il concetto BE(CAT), UNDER(TABLE) è utile se l'interlocutore non conosce il contesto o se noi siamo con lui in cucina, altrimenti avremmo potuto preferire la specificazione *kitchen table*, *tavolo da cucina*. La formulazione della frase è per la gran parte guidata lessicalmente (Ferreira, 2010; Bock, Ferreira, 2014**).** Per alcuni teorici la formulazione attinge a conoscenze di tipo procedurale e dichiarativo (Levelt, 1989), mentre per altri esse sono distinte (Chang, Fitz, 2014). La conoscenza dichiarativa di una parola include che la parola *gatto* è un nome e quella procedurale include il processo con cui si costruisce la forma della parola e si inserisce un nome in una frase. Nel modello di Levelt *et al*. (1999), il lemma è una particolare rappresentazione della memoria dichiarativa, collegato con rappresentazioni concettuali, sintattiche e fonologiche della parola. Il lemma della parola *gatto* lega informazioni concettuali come il concetto lessicale GATTO, la sua connessione con altri concetti come PELOSO, MIAGOLANTE, ANIMALE etc. La rete lessicale consiste di concetti lessicali (GATTO, CANE, CASA, PADRONE) ed è rappresentata nella corteccia temporale

anteriore–ventrale. Il lemma nella sezione mediana del giro temporale mediano sinistro (MTG) dà input ai fonemi e dà input e output alle forme lessicali e ai morfemi nel giro superiore temporale anteriore sinistro e nel MGT posteriore, cioè l'area di Wernicke, poi dà output ai fonemi nel giro frontale inferiore posteriore sinistro, cioè l'area di Broca, e al programma motorio delle sillabe nel giro ventrale precentrale. Le localizzazioni si fondano su note metanalisi (Indefrey, Levelt, 2004; Indefrey, 2011) integrate con dati provenienti dal campo dell'afasia.

I concetti lessicali farebbero parte di un *hub* di rappresentazioni concettuali sovramodali, elaborate nel lobo temporale anteriore, che integrano caratteri (*features*) specifiche per modalità che sono rappresentate in aree percettive e motorie diffuse nel cervello (l'altro nostro saggio in questo volume per una discussione critica al §3). I lemmi specificano le proprietà grammaticali delle parole rappresentate nel STG posteriore sinistro e nel MTG. Le proprietà grammaticali sono diffuse all'area di Broca per la codifica dei sintagmi e delle frasi.

3. Interferenze semantiche e relazioni categoriali

Roelofs (2018) rielabora il modello Weaver ++ per fornire un resoconto computazionale unitario della selezione lessicale relativamente a tre paradigmi utilizzati: studio della *cumulative semantic interference* nel *continuous naming*, della interferenza semantica nel *blocked–cyclic naming*, della interferenza semantica con parole distrattore nella *picture word interference*. Questi paradigmi studiano l'interferenza semantica a partire dalla relazione dell'esemplare visualizzato nella figura da denominare con un esemplare a lui vicino per appartenenza ad una catego-

ria superiore. Il compito richiede di selezionare la parola *forchetta* nel contesto di *bicchiere* o di *cucchiaio*. L'accuratezza e la velocità sono minori nei contesti di relazione tra immagine e categoria e maggiori in quelli in cui non c'è relazione.

Nel *continuous naming* (Belke 2013, Howard *et al.* 2006) i soggetti devono denominare figure che provengono da varie categorie (*stoviglie*, *uccelli*, *attrezzi da lavoro*). Le immagini non vengono ripetute, ma ci sono si solito cinque esemplari per ogni categoria (*forchetta*, *cucchiaio*, *bicchiere*, *sega*, *cacciavite*). Nei vari *trials* dell'esperimento le immagini di differenti categorie sono mischiate fra loro, variando in tal modo il numero di immagini non connesse che occorrono tra i membri della categoria.

Il compito di denominazione della figura viene valutato a partire da ogni posizione ordinale all'interno della categoria. Per esempio in una sequenza possiamo avere l'immagine di una *forchetta* seguita da un numero di immagini non collegate e da un *cucchiaio*, che invece è collegata. Si dice allora che la forchetta ha la prima posizione all'interno della categoria delle *stoviglie* e *cucchiaio* ha la seconda posizione. La risposta di denominazione aumenta linearmente con la posizione dell'esemplare da nominare. L'effetto di interferenza semantica cumulativa è indipendente dal numero esatto di altri *items* irrelati che occorrono tra ciascun esemplare della categoria se il *range* di esemplari si mantiene tra 1 e 8. L'effetto scompare quando il numero di immagini irrelate è più alto di otto.

Nell'esperimento del *blocked–cyclic naming* (Harvey, Schnur 2016) i soggetti ripetono per sei volte la denominazione di un piccolo insieme di immagini (spesso cinque). In un ciclo omogeneo, le immagini appartengono tutte alla stessa categoria e il compito consiste in "nomina

una forchetta, un bicchiere, un cucchiaio, una tazza e un coltello". In un ciclo eterogeneo le immagini provengono da differenti categorie e il compito consiste in "nomina una forchetta, un cigno, una scure, un letto e una bocca". Le immagini sono le stesse in entrambi i cicli, ma varia il modo in cui sono raggruppate. L'effetto di blocco si misura all'interno delle categorie. Il tempo di risposta di denominazione della figura è maggiore nella condizione del ciclo omogeneo che in quello eterogeneo. L'interferenza non è presente nel primo ciclo, ma può aumentare nei cicli successivi, sebbene non si riscontri l'effetto nei soggetti sani. Nel *picture word interference* si denominano le figure cercando di ignorare distrattori parlati o scritti come ad esempio dover dire il nome della forchetta mostrata con sopra scritta la parola connessa *bicchiere* o quella non connessa *cigno*.

Questi paradigmi partono dal presupposto che gli effetti di interferenza semantica nel compito di denominazione di immagini siano legati alla produzione del parlato (Levelt 1989, Roelofs 2014) dal momento che in questo processo il parlante seleziona le parole dalle informazioni concettuali della memoria semantica, e l'interlocutore recupera questa informazione concettuale a partire dalla parola.

I vecchi modelli computazionali che erano stati applicati alla interferenza immagine–parola non riguardavano le scoperte di questi tre nuovi paradigmi. Al contempo i modelli computazionali applicati agli effetti *semantico–cumulativi* e *semantic blocking* (Howard *et al.* 2006, Oppenheim *et al.* 2010) non hanno riguardato l'interferenza nel compito di denominazione dell'immagine. Roelofs (2018) respinge così le critiche al modello WEAVER ++, che sarebbe falsificato dagli effetti semantico–cumulativi, accettando però l'invito a modificarlo lanciato da Howard. Si mantengono due assunti chiave del modello del lemma:

1. La selezione lessicale avviene per competizione, 2. l'origine degli effetti semantici si trova al livello concettuale e si localizza nella fase di selezione lessicale.

Howard pensa che i nodi del concetto unitario sono legati con i nodi lessicali da connessioni eccitatorie in avanti. L'immagine di *forchetta* attiva il nodo del concetto corrispondente (FORCHETTA) e, parzialmente, i nodi del concetto delle parole semanticamente connesse (CUCCHIAIO, BICCHIERE, COLTELLO, TAZZA) e l'attivazione si diffonde dai concetti ai nodi lessicali. Per ogni *step* l'attivazione di ogni nodo è aggiornata da una funzione di attivazione standard. Quando l'attivazione del nodo lessicale *forchetta* supera la soglia, il nodo viene selezionato e la connessione tra nodo lessicale *forchetta* e il concetto FORCHETTA viene rafforzata. Quando, in un secondo momento, un'altra immagine della stessa categoria deve essere denominata, mettiamo che essa sia quella di *cucchiaio*, il nodo lessicale della precedente categoria (cioè *forchetta*) è più fortemente attivato a causa del fatto che l'attivazione è stata rafforzata. L'inibizione dall'item lessicale precedente sarà più forte e la selezione del *target cucchiaio* sarà più dilazionata nei termini di maggiori passaggi richiesti per raggiungere la soglia di attivazione. L'inibizione aumenta col numero di immagini da nominare in una particolare categoria, producendo l'interferenza cumulativa semantica. Dal momento che il rafforzamento delle connessioni si estende nel tempo, l'interferenza è indipendente dallo sfasamento o dal ritardo tra le immagini.

Oppenheim assume che i nodi delle proprietà concettuali–semantiche sono legati ai nodi lessicali da connessioni eccitatorie e inibitorie *feedforward–only*. In questo modo ESSERE FATTO DI METALLO e AVERE DEI DENTI sono legati in modo eccitatorio con forchetta, mentre alcune di queste proprietà come AVERE DEI

DENTI sono legate in modo inibitorio a *cucchiaio*. Vi è un previo apprendimento della rete per cui, per ogni parola come *forchetta*, viene settata a 1 l'attivazione delle proprietà e del nodo lessicale pertinente, e a 0 le attivazioni di altre parole. La simulazione al computer fatta da Oppenheim e colleghi dell'esperimento di Howard *et al.* (2006) conferma le tesi di questi ultimi sul fatto che l'inibizione aumenta col numero di immagini da nominare in una particolare categoria.

Roelofs propone che la disponibilità, dopo la selezione, del concetto lessicale produce un *bias*, una propensione, un pregiudizio verso il concetto nel *processing* successivo di un altro concetto lessicale all'interno dello stesso contesto di categoria semantica o tematica. Nominare la figura di una forchetta fa propendere verso il concetto collegato con la immagine della forchetta nella denominazione successiva della immagine semanticamente relata, quella di un *cucchiaio* per esempio. L'inserimento di un *bias* temporale di tipo concettuale spiega meglio i fenomeni di interferenza rispetto alle spiegazioni nei termini di apprendimento dalla durata consistente di pesi relativi alla connessione concettuale–lessicale.

4. Il problema della *convergenza* e l'approccio non componenziale

4.1. *Hypernym problem*

Il problema dell'iperonimo (*hypernym problem*), o della ereditarietà dei tratti, viene segnalato per primo da Levelt (1989, 201–214) come uno dei problemi principali e spinosi, un banco di prova (*touchstone*) della teoria dell'accesso lessicale. Si tratta più in generale del proble-

ma della *convergenza* del parlante verso la singola e unica parola che egli vuole esprimere, relativa soprattutto alla selezione da parte dei soggetti del peculiare livello di astrazione pertinente alle loro intenzioni comunicative e al contesto in cui viene enunciato il messaggio. Per cui, ad esempio, diciamo "Quel *leone*", e non semplicemente "Quell'ANIMALE", o "Mio *padre*", e non semplicemente "Il mio GENITORE".

Levelt (1989, 213) definisce così l'*hypernym problem*: Quando il significato di un lemma A implica il significato di un lemma B, B è un iperonimo di A. Se le condizioni concettuali di A sono soddisfatte, allora lo sono anche necessariamente quelle di B. Dunque, se A è il lemma corretto, B sarà anche esso stesso recuperato".

Se il parlante vuole esprimere il concetto *cane* o *leone*, allora saranno soddisfatte le condizioni di ANIMALE. Il parlante però non seleziona l'iperonimo ANIMALE al posto dell'iponimo *cane* o *leone*. Il nostro linguaggio non è fatto solo di iperonimi, ma si situa su livelli gerarchici più specifici e pertinenti impedendo una ereditarietà dei tratti per *default*. Se un adulto dice al bambino la parola *tigre* per indicare l'animale che sta dentro la gabbia dello zoo o al circo, non per questo intenderà trasferire nella sua selezione lessicale tutti i tratti categoriali della definizione di tigre e cioè che essa è un ANIMALE che oltre ad *avere quattro zampe*, *avere il manto a strisce*, *avere una lunga coda*, è anche *particolarmente feroce e carnivoro*. La selezione lessicale, come vedremo sotto, sembra più sintetica e olistica e meno analitico–componenziale (o composizionale) della definizione concettuale. Inoltre essa non rispecchia la transitività categoriale ovvero la proprietà, da cui sorge l'*hypernym problem*, per cui se il concetto A è iperonimo di B e se il concetto B è iperonimo di C, allora C sarà iponimo anche del concetto A. Questa proprietà tran-

sitiva è uno dei capisaldi della normatività logica delle relazioni fra concetti. Come vedremo, per risolvere il problema della eredità dei tratti occorre ripensare al concetto lessicale come a qualcosa di non scomponibile e quindi di intransitivo (Roelofs 1997, 2014). Sappiamo già che, in generale, ci sono casi di intransività categoriale anche nel sistema astratto normativo. Hampton (1982) analizza il caso delle parole composte come *ski–lift* e *car–seat*, cui possiamo aggiungere *sedia elettrica*, per mostrare come, pur appartenendo al subordinato *sedia*, questi concetti non appartengano al sovraordinato MOBILE. Questa intransitività si ritrova tipicamente nei concetti congiuntivi o disgiuntivi complessi, di cui i subordinati di MOBILE sono un esempio, ma che possiamo vedere anche in veste di sovraordinato intermedio come nel caso de *gli sport che sono anche dei giochi* o de *gli sport che non sono anche dei giochi*. Sappiamo infatti che ci sono giochi che non sono anche sport (*scacchi*) e viceversa sport che non sono anche giochi (*pugilato*, *lancio del peso*). Questi ultimi quindi non rientrerebbero nelle macrocategorie SPORT o GIOCO per pura applicazione della proprietà transitiva. Per questo motivo Roelofs (1997) cercherà di superare l'*hypernym problem* con un approccio non componenziale e radicalmente olistico.

È vero altresì che i soggetti normali possono intenzionare liberamente, nel loro atto di parole, tutta una serie di richiami di componenti e proprietà (*features*) dell'esemplare e utilizzare con disinvoltura gli iperonimi quando non sanno come specificare il concetto, nei casi in cui la vaghezza è necessaria o per varie ragioni semantico–pragmatiche (Velardi 2005, 2009). Infatti "questi attributi possono rientrare nella nostra focalizzazione categoriale, ma la loro selezione dipende da cosa c'è nella mente del parlante e non da un processo logico, astratto ed estrin-

seco di passaggi di proprietà" (Velardi 2005, 238). La mente non è sottoposta alla sequenzialità analitica di questi passaggi, come lo sarebbe un calcolatore in preda al processo sequenziale di calcolo della eredità dei tratti. La prospettiva radicale di Roelofs supera questa astratezza e l'*output* lessicale appare come un composto olistico, non scomponibile in tratti, con un blindamento semantico tale per cui la parola esprime solo l'oggetto utile al contesto di enunciazione, senza nemmeno dover contenere tratti utili, perché appunto non ha natura componenziale. Certamente, collocandosi all'estremo opposto della teoria componenziale, anche questa impermeabilità ai tratti suscita perplessità e appare un modo radicale di superare l'*hypernym problem*. La fissazione olistica di una selezione lessicale univoca, che stabilisce in maniera radicale ed esclusiva il livello di astrazione dentro cui si colloca il parlante, sebbene sia una soluzione elegante, e per certi versi giusta, d'altra parte non sembra rispecchiare appieno da una parte la complessità della semantica dell'enunciazione con le comunicazioni dei livelli di astrazione, dall'altra la connessione non sempre trasparente tra *conceptualization problem* e *verbalization problem.* Spesso la verbalizzazione dipende dalle parole semplici o composte che il dizionario di una lingua mette a disposizione rivelando quanto sia utile la scomponibilità. Se io in inglese voglio dire *cavallo femmina* posso usare la parola singola mare, se invece voglio dire *elefantessa* ho un *verbalization problem* più stringente e sono costretto a utilizzare la locuzione *female elephant* (Levelt *et al.*, 1999). Così la frase, citata già in §2, *il gatto è sotto la tavola* esprime il concetto BE(CAT), UNDER(TABLE), ma l'interlocutore può usare *tavola* se l'interlocutore non conosce il contesto o se noi siamo con lui in cucina, altrimenti si sarebbe potuto preferire la specificazione *kitchen table*, *tavolo da cucina* (Roelofs, Fer-

reira, *in press*). Nessuno impedisce che il parlante possa riferirsi mentalmente al *tavolo da cucina* dicendo *tavolo* e così, in generale, riferirsi con un termine generale a qualcosa di più specifico. A prescindere dai fini comunicativi è sempre possibile al parlante ricorrere indifferentemente all'iponimo o all'iperonimo per designare un oggetto, compiendo quello che possiamo chiamare *shifting* categoriale, sfruttando appieno la comunicazione tra livelli di astrazione e legando creativamente *conceptualization problem* e *verbalization problem*.

Sappiamo inoltre come i soggetti patologici suppliscano all'accesso al lemma specifico utilizzando con efficacia l'iperonimo superiore. Questo accade per esempio nell'afasia o nel caso della demenza semantica, che ha un regresso gerarchico, in cui permane forte l'uso dei sovraordinati (Patterson et *al.*, 2007*;* Lambon Ralph *et al*., 2010*).*

La strada del modello del lemma sembra andare però verso una radicale specificità e non scomponibilità. Oltre all'approccio olistico, che approfondiamo in §3.1, Levelt (1989, 211–212) attribuiva grande importanza al principio connessionista della distribuzione e della *diffusione dell'attivazione* per spiegare gli aspetti paralleli dell'accesso lessicale. Roelofs (1992, 1997, 2018) fornisce un modello maturo basato su questa indicazione attraverso cui cercare di risolvere il problema dell'iperonimo. Levelt (1989, 213) aveva cercato di risolvere l'*hypernym problem* attraverso tre principi teorici: lo *uniqueness principle*, il *core principle*, il principio di specificità. La nozione di *core meaning* indica che il lemma possiede un core concettuale privilegiato rispetto a quello di altri lemmi concorrenti. Questa salienza e sensitività contestuale sono espresse dai test della negazione usati con verbi di locomozione in cui i soggetti devono completare frasi del tipo *They do*

not ski, but they ... Nessuno risponderebbe con i verbi *breathe* o *think*, ma quasi tutti completano con *skate*, che viene considerato più appropriato al contesto di una azione fatta su una superficie ghiacciata. Il *core* è relativo a questo significato condiviso dai verbi *ski* e *skate*, che costituisce una parte non mutuabile da altri verbi, chiamata presupposizione del lemma. Per lo *uniqueness principle* non ci "sono due *item* lessicali che possiedono lo stesso *core meaning*". Il *core principle* aggiunge che "un item lessicale è recuperato solo se le condizioni del suo core sono soddisfatte dal concetto che deve essere espresso" (Levelt 1989, 213). I verbi *ski* e *skate* hanno come iperonimo il verbo GLIDE (SCIVOLARE), ma se il parlante vuole riferirsi all'azione con il verbo GLIDE, allora il *core principle* non gli permetterà di selezionare il lemma per *ski*. Può ancora succedere però che, se il parlante vuole selezionare il verbo *ski*, può recuperare GLIDE in quanto il significato di *ski* implica GLIDE. Per evitare questo scenario, Levelt aggiunge il principio di *specificità*: "di tutti gli item di cui il concetto soddisfa le condizioni legate al core, viene recuperato solo il più specifico" (ibid.). Roelofs (1992) prosegue su questa strada mettendo a frutto la nozione di *diffusione dell'attivazione* integrata però da quella di *non composizionalità* del concetto lessicale che abbiamo visto essere radicale (Roelofs 1997). Per fare questo egli cerca di mostrare i limiti delle teorie componenziali (Jackendoff 1983, 1987, 1990), assai egemoni nel Novecento, e i pregi di un approccio al concetto e alla parola considerati come un intero olistico. Una strada che, come abbiamo visto sopra in §3, ha mantenuto anche nelle simulazioni più recenti (Roelofs 2014, 2018). Approfondiamo dunque la strategia con cui Roelofs difende l'approccio olistico.

4.2. Oltre le teorie componenziali del concetto lessicale

L'approccio componenziale è tipico delle semantiche a tratti in cui si pensa che concetti e/o parole sia espressi da componenti o *features* a vari livelli di generalità. Essi spiegano i deficit specifici per categoria dei pazienti neuropsicologici ricorrendo a questi tratti che sarebbero rappresentati "in forma funzionalmente distinta" (Basso, Chialant 1992, 107). Per questo un soggetto può avere difficoltà a riconoscere o denominare oggetti caratterizzati dalla presenza del tratto ANIMATO o INANIMATO. Il termine *mela* è scomponibile nei tratti: FRUTTO E COMMESTIBILE E ROTONDO E CON LA BUCCIA LISCIA E ROSSO E GIALLO (ivi, 98). Nell'approccio componenziale le parole vengono recuperate in memoria tramite una *combinazione di concetti primitivi* o comunque riducibili a primitivi semantici fondamentali. Secondo questo modello il lemma *padre* ha una rappresentazione del tipo (MASCHIO (X), GENITORE (X, Y). Il lemma *madre* una rappresentazione del tipo (FEMMINA (X), GENITORE (X, Y). Gli elementi si coimplicano nell'attività di recupero di più parole. La rappresentazione astratta GENITORE (X, Y) è coinvolta nel recupero di *madre*, *padre* e di *genitore* stesso. Le semantiche composizionali sono afflitte dal problema della inclusione dei tratti. Se a guidarlo è una logica binaria, allora non possiamo inserire contemporaneamente i tratti ROSSO E GIALLO per la *mela*, come nella definizione fornita da Basso e Chialant (1992). Se però scegliamo solo uno di questi colori, allora non diamo ragione del fatto che esistano mele sia di colore ROSSO che GIALLO. L'assunzione implicita o esplicita, di un certo modello semantico comporta dei problemi che normalmente negli studi di psico o neurolinguistica non vengono presi in con-

siderazione nella loro ampiezza e che invece, hanno una loro complessità teorica (Perconti 2001, 183). Uno di questi casi è proprio il problema della composizionalità del significato lessicale. È merito di Roelofs (1997) aver esplorato nel dettaglio il problema proponendo soluzioni che certamente rimangono ancora da verificare in sede linguistico–cognitiva teorica.

Per le teorie non componenziali (Fodor 1976, Fodor *et al.* 1975; Fodor *et al.* 1980; Roelofs 1997, 2014, 2018), come anche quella della *diffusione della attivazione* (Collins, Loftus 1975), una rappresentazione astratta olistica e sintetica del concetto elicita il recupero di una sola parola presa come un intero. La rappresentazione di PADRE (X, Y) innesca il recupero del lemma *padre* e solo in seguito la memoria semantica specifica, al di fuori del messaggio, le proprietà di (MASCHIO (X) e GENITORE (X, Y) presenti nella definizione del concetto della parola. È un errore, compiuto per esempio da Jackendoff, pensare che l'analisi componenziale del concetto si debba sovrapporre alla scomposizione lessicale della parola. Al contrario, come in Fodor, convivono due modelli differenziati, analitico per i concetti e olistico per le parole.

Roelofs (1997, 39) disegna le sue simulazioni quando prevalevano ancora i modelli amodali dei concetti. Cita quindi concetti astratti come DEMOCRAZIA, PACE, e quelli di NUBILE, MADRE, UCCIDERE o CAUSA e MORIRE (Levelt 1989, 93) per mostrare come non siano esprimibili attraverso modalità sensoriali. Oggi sappiamo che vi è un paradigma di ricerca *embodied* che mostra l'esistenza di una qualche base sensorimotoria perfino per i concetti astratti (Borghi, Binkofski 2014)[2]. È possibile

[2] Si veda l'altro saggio pubblicato in questo volume sulla relazione e il conflitto tra rappresentazione e *embodiment*.

che questo approccio, o l'approccio a due stadi, multimodale + amodale che proponiamo nell'altro saggio di questo volume, possano dare una nuova soluzione ai problemi posti da Roelofs, ma preferiamo non occuparci di questo tema nel presente contributo. Osserviamo solo che il modello del lemma, conduce ad una riconsiderazione della unitarietà e sinteticità del concetto, vicina alla nostra esperienza sensorimotoria degli oggetti, legando aspetti dichiarativi e procedurali in una visione che rimane però più amodale (Roelofs, Ferreira *in press*, 5–7). Occorre dunque inglobare le informazioni analogiche visive nella rappresentazione astratta.

Come gli oggetti siano codificati nella memoria e siano riconosciuti, resta un problema sia per l'approccio componenziale che per quello olistico. I primitivi discreti, proposti da Jackendoff (1990), sono però inadeguati perché, per disambiguare entrate lessicali come *anitra* e *oca*, occorre introdurre una immagine visiva della loro differenza, ma, da una parte esprimere questa differenza nei termini di componenti binarie come +– *long neck* è impraticabile, dall'altra nel modello componenziale, le immagini mentali visive restano invisibili alla rete. Non lo sono invece per un approccio olistico in cui l'informazione sulla forma resta invisibile solo per la codifica sintattica del messaggio, ma è associata in memoria alla componente concettuale ai fini del recupero lessicale, senza essere direttamente coinvolta nel processo di decisione lessicale sulla produzione del messaggio più pertinente.

Fodor (1998) critica l'idea che i concetti siano un insieme di caratteristiche definitorie, mostrando che essa poggia su una nozione di definizione che si basa a sua volta su nozioni problematiche come quella di analiticità. La struttura dei concetti non è quella delle definizioni. Roelofs si smarca dalla sovrapposizione, generatrice di confu-

sione, tra parola lessicale, concetto da definire e concetto da acquisire. La componenzialità tende alla definizione teorica di un termine. Per esempio cane è un *quadrupede* e *abbaia*, laddove invece per il recupero lessicale esso è solamente il *cane* e questo accesso non richiede necessariamente un *processing* degli attributi del concetto, ma fa di CANE (X) una parte compatta del messaggio per cui la parola *cane* viene recuperata direttamente, in modo olistico, come un intero.

Fodor (1981) prevede che i concetti lessicali siano prodotti dalla combinazione di concetti primitivi innati. L'approccio olistico avrebbe dunque difficoltà a spiegare l'apprendimento delle parole. Ma un primitivo computazionale non deve per forza coincidere con un *developmental primitive* o un *definitional primitive*. Il concetto di NUBILE è composto da NON SPOSATO, UMANO, ADULTO, FEMMINA. FEMMINA ricorre nella definizione di molteplici concetti senza essere definito, in qualità di *primitivo definizionale*. Ma non per questo deve ricorrere in tutti gli accessi lessicali di tutte le parole nella cui definizione è presente. Se un bambino usa il *definitional primitive* FEMMINA per apprendere concetti come ZIA, MADRE, PRINCIPESSA, STREGA, allora esso diventa anche un developmental primitive. Un approccio olistico risolve la questione perché non esclude che componenti composizionali possano entrare a fare parte sia dell'acquisizione che della rappresentazione del concetto, ma esclude che essi abbiano una relazione con l'accesso lessicale alla parola.

L'approccio olistico risolverebbe anche altri problemi. Infatti i parlanti non esibiscono una conoscenza precisa delle proprietà degli iponimi. Come osservato da Putnam (1975) essi hanno uno stereotipo molto generico dei generi naturali, per esempio di alberi come *larice* o *faggio*. Una

teoria olistica richiede che il soggetto possiede almeno una proprietà distintiva nella memoria semantica, mentre un approccio non composizionale richiede più semplicemente che il concetto LARICE(X) sia legato alla specificazione IS–A(LARCHX), ALBERO(X).

Un altro caso interessante è quello del colore come *rosso* che non può rientrare come componente nella rappresentazione di COLORE(X), altrimenti quando dobbiamo rappresentare ROSSO(X) dovremmo inserire pure *rosso* come suo proprio componente (Fodor *et al.* 1980).

4.3. *Una soluzione non componenziale dell'hypernymy problem*

Il campo principale in cui i modelli componenziali falliscono è sicuramente quello della convergenza dell'accesso lessicale, del convergere cioè da parte del parlante verso la parola appropriata e pertinente, quella più adeguata alle sue intenzioni enunciative e al contesto nel quale si trova a emettere il messaggio. Oltre all'*hypernymy problem*, troviamo la *word–to–phrase synonymy* che pone problemi simili dal momento che sia la frase "è un padre" sia la frase "è un genitore maschio" hanno una sola medesima rappresentazione (Fodor 1976). Nel modello componenziale il processo di codifica sintattica costruisce una frase elaborando che NUBILE è una rappresentazione composta dai tratti NON SPOSATO, UMANO, ADULTO, FEMMINA; che CANE è QADRUPEDE ABBAIANTE, che UCCIDERE è CAUSARE LA MORTE. Come nota Levelt (1989, 93) il parlante sa implicitamente tutte queste cose, ma non le esplicita nel come "parti del messaggio preverbale". Per una teoria olistica questa ridondanza scompare. Rimane però che una frase come "è nubile" e una frase come "è non sposata, umana, adulta, femmina" hanno la

stessa rappresentazione soggiacente e che occorre spiegare perché il parlante sceglie una verbalizzazione singola nel primo caso o una verbalizzazione sintagmatica nominale nel secondo. Le teorie componenziali non spiegano come il recupero della parola converga verso il lemma rilevante. La strategia di Roelofs consiste invece nel considerare le singole parole del linguaggio come concetti indipendenti.

Torniamo al problema dell'iperonimo. Il concetto di *cane* è rappresentato dalle componenti ABBAIARE(X), MUOVERSI(X), VIVERE(X), QUADRUPEDE(X), le quali a loro volta sono comuni al suo iperonimo ANIMALE(X). Per questo la rappresentazione concettuale dell'iponimo contiene tutti i componenti dell'iperonimo più quei componenti che lo differenziano da un possibile co–iponimo. I *canidi* differiscono così dai felini, anche se con la difficoltà dell'inserimento di *gatto* che fa eccezione, per componenti altamente differenziali come RUGGIRE(X) e ANIMALE SELVATICO (X). Differiscono dagli uccelli per le proprietà BIPEDE(X) e FLUTING(X). Al contrario, in un approccio non componenziale, le parole *padre* e GENITORE, *cane* e ANIMALE sono recuperate dalla memoria come elementi autonomi del vocabolario: CANE(X) E AMIMALE(X).

La teoria olistica confina il problema della *convergenza* distinguendo tra l'*hypernym problem* per il recupero della parola dalla memoria e un *superordinate problem* per la codifica del messaggio. Se una persona vuole riferirsi alla propria madre, nel senso di concettualizzarla come MADRE e non come GENITORE, selezionando il subordinato troverà le condizioni concettuali del sovraordinato. Per la non componenzialità GENITORE (X,Y) (MASCHIO (X) PADRE (X, Y) sono primitivi concettuali e sono parti del messaggio da recuperare come parole a parte. Ma non basta. Non solo ogni concetto lessicale corrisponde ad un

nodo olistico della rete, ma in Roelofs (1992) la *diffusione dell'attivazione*, in contrasto con Dell (1986) e in accordo con Collins, Loftus (1975), possiede a) un livello concettuale con dei nodi e dei *link* concettuali, b) uno livello sintattico con nodi e *link* di lemma, c) uno livello di forma di parole con nodi e *link* di lessema *input* (di tipo ortografico) e d) di lessema *output* (sulle proprietà morfo–fonologiche della parola parlata). Per risolvere i due problemi della *convergenza* si assume che i nodi concettuali sono solo indirettamente collegati ai nodi dei lemmi attraverso una rappresentazione concettuale non componenziale. Il lemma viene recuperato dal rafforzamento del livello di attivazione del nodo del concetto che deve essere verbalizzato. L'attivazione si propaga per il livello sintattico e il nodo del lemma attivato più in alto nella rete viene selezionato. Nel nominare un dato oggetto sono coinvolti quattro stadi: 1. l'identificazione concettuale dello stimolo come *input* percettivo porta alla identificazione dell'oggetto, 2. recupero del lemma, 3. codifica della forma della parola, 4. articolazione fonica del nome dell'oggetto. Il modello di Roelofs focalizza solo lo stadio del lemma recuperato a partire da un concetto che si deve verbalizzare. La categorizzazione di un oggetto (*un cane è un ANIMALE*) richiede il recupero del sovraordinato corrispondente. Il flusso di informazione tra identificazione concettuale e recupero del lemma è continuo. Il nodo del lemma contiene informazioni sul senso della parola, sulle sue proprietà sintattiche e sui suoi lessemi di *input* e di *output*. In una visione non componenziale, ogni nodo rappresenta un singolo concetto come CANE, ANIMALE, ABBAIO. I *link* tra i nodi consistono in *labelled pointers* che esprimono la relazione tra due concetti. Per esempio il *link* IS–A indica che CANE è un sottotipo di ANIMALE e il *link* CAN indica che CANE può ABBAIARE. I *link* differiscono per accessibilità e

questa dipende dal peso stabilito dalla *spreading activation*. L'identificazione concettuale di un oggetto riguarda la rappresentazione su vari livelli di astrazione, preferibilmente il livello degli oggetti di base (Rosch *et al.* 1976). Per questo ci sono anche nodi per la forma visiva degli oggetti relativi soprattutto alla identificazione di oggetti basata sulla forma. Nella verbalizzazione di CANE viene selezionato il nodo CANE che eccede la soglia rispetto ad altri nodi che competono, come ANIMALE o CANIDE, che sono coattivati, ma ricevono una piccola proporzione di attivazione rispetto a CANE. Nella verbalizzazione di MADRE viene selezionato il nodo MADRE(X,Y) che eccede la soglia rispetto ad altri nodi che competono come GENITORE(X,Y) che è coattivato ma riceve una piccola proporzione di attivazione rispetto a MADRE. In questo modo è garantita una ed una sola uscita lessicale attraverso un approccio olistico radicale che supera parecchi problemi dell'approccio componenziale, con tutti i limiti di rispecchiamento della reale psicologia della categorizzazione e dell'atto linguistico che abbiamo segnalato in §4.1 e che occorrerà approfondire nel dettaglio in future ricerche.

5. Conclusioni

I risultati della nostra indagine sulla teoria dell'accesso lessicale, vista dalla prospettiva di uno dei modelli più efficaci per il suo trattamento e cioè il modello di Levelt *et al.* (1999) rivisitato da Roelofs (1992), hanno mostrato che vi sono due importanti problemi teorici. Il primo riguarda il rapporto che intercorre tra le rappresentazioni categoriali degli oggetti e degli esemplari cui le parole rinviano, con i loro livelli di astrazioni e i rapporti presenti dentro questi livelli che sono governati da regole logiche come la transi-

tività e la ereditarietà dei tratti, e la convergenza dei parlanti verso la parola più pertinente e adeguata a indicare che tipo di rappresentazione ha in mente il parlante e a che tipo di referente si riferisce rispetto ai vari livelli di astrazione categoriale entro i quali egli può collocare l'*output* lessematico della propria rappresentazione (§4.1). Questo problema è chiamato problema dell'iperonimo (*hypernym problem)* o problema della *ereditarietà dei tratti* o in generale *problema della convergenza.*

Il secondo problema riguarda il modo in cui l'accesso lessicale può superare l'ostacolo della *ereditarietà (§§4.2,4.3*.). Levelt (1989) ha proposto di aggiungere dei *principi vincolanti come* lo *uniqueness principle*, il *core principle*, il principio di *specificità* di cui abbiamo abbondantemente parlato in §4.1. Roelofs ha invece intrapreso una revisione della teoria composizionale, o componenziale, dei concetti e delle parole, optando per un modello non analitico, sintetico–olistico dell'accesso lessicale alle parole fonologiche (§§4.2, 4.3) mantenendo aperta la possibilità che la composizionalità possa influenzare il *processing* meramente concettuale delle rappresentazioni astratte connesse ai significati delle parole, ma non il *processing* relativo alla produzione linguistica con la selezione del lessema pertinente e la verbalizzazione dei concetti stessi. Da qui il rifiuto della proposta di Jackendoff di assimilare l'analisi componenziale del concetto alla scomposizione lessicale della parola e la preferenza per l'approccio moderato di Fodor che prevede che un approccio analitico per i concetti possa convivere con un approccio olistico per le parole.

Attraverso una indagine teorica sui modi in cui il parlante può intenzionare la pertinenza di una parola rispetto alla rappresentazione categoriale di quest'ultima, il presente lavoro mostra che la comunicazione tra i livelli di

astrazione delle categorie e delle parole di riferimento (sovraordinato, basico, subordinato, esemplare individuale) sono complessi e il problema della ereditarietà dei tratti, o della iperonimia, è solo uno degli aspetti che rivelano questa complessità.

Si mostra inoltre come sia difficile separare chiaramente e radicalmente la dimensione analitica da quella olistica né per quanto riguarda il *processing* meramente concettuale e rappresentazionale astratto, né per quanto riguarda il *processing* della produzione linguistica e della *convergenza* della verbalizzazione verso il lessema che risulta più pertinente e adeguato al contesto. La analiticità si intreccia profondamente con la non analiticità e l'olismo della selezione lessicale non esclude questo intreccio all'interno del processo rappresentazionale e del legame tra rappresentazioni mentali e *output* lessematico del parlante. I nessi tra concetto e parola rimangono complessi. Il parlante può ricorrere in maniera molto elastica all'iponimo o all'iperonimo per designare un oggetto, compiendo quello che possiamo chiamare *shifting* categoriale, sfruttando la comunicazione tra livelli di astrazione che lilega nella sua mente e nel linguaggio intrecciando creativamente il *conceptualization problem* con il *verbalization problem*. Pur essendo un geniale tentativo di approcciare i processi della produzione linguistica, l'olismo lessicale di Roelofs, mantenuto e non smentito per tutto lo sviluppo del modello del lemma dal 1992 fino alle più recenti pubblicazioni (Roleofs 2014, 2018; Roelofs, Ferreira *in press*), sembra più ripresentare che risolvere il grande enigma dell'accesso lessicale e della ereditarietà dei tratti.

Questo fatto mostra come l'analisi linguistico-concettuale teorica possa ancora dare notevoli contributi ai modelli della produzione linguistica presenti nel panorama psico–linguistico, computazionale, neuro–psicologico,

stimolando gli studiosi ad una maggiore considerazione della complessità del fenomeno della selezione lessicale e della teoria del significato.

Riferimenti bibliografici

ABRAMS L. (2008). *Tip–of–the–Tongue States Yield Language Insights*, «American Scientist*», 94(3).*

BASSO A., CHIALANT D.*, I disturbi lessicali nell'afasia: proposte di riabilitazione*, Milano, Masson, *1992.*

BELKE E. (2013)*, Long–lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical–semantic encoding*, «Journal of Memory and Language», 69, 228–256.

BIEDERMANN B., RUH N., NICKELS L., COLTHEART M., (2008), *Information retrieval in tip of the tongue states: new data and methodological advances*, «Journal of Psycholinguist. Res.», 37, 171–198.

BOCK K., FERREIRA V. S., (2014), *Syntactically speaking*, in GOLDRICK M., FERREIRA, V. S., MIOZZO M., (eds.), *The Oxford Handbook of Language Production*, Oxford University Press, Oxford.

BORGHI A.M., BINKOFSKI F., (2014), *Words as social tools: an embodied view on abstract concepts*, Springer, New York/ Berlin.

CARAMAZZA A. (1997). *How many levels of processing are there in lexical access?,* «Cognitive Neuropsychology»*, 14: 177–208.*

CARAMAZZA A., MIOZZO M., (1997)*, The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip–of–the–tongue' phenomenon*, «Cognition», 64, 309–343.

CHANG F., FITZ H., (2014), *Computational models of sentence production: A dual–path approach*, in GOLDRICK M., FERREIRA V. S., MIOZZO M., (eds.), *The Oxford Handbook of Language Production*, Oxford University Press, New York.

COLLINS A. M., QUILLIAN M.R., (1969) *Retrieval time from semantic memory*, «Journal of Verbal Learning and Verbal Behavior», 8:241–248.

COLLINS A. M., LOFTUS, E.F. (1975), *A spreading–activation theory of semantic processing*, «Psychological Review», 82(6), 407–428.

DELL G. S. (1986), *A spreading–activation theory of retrieval in sentence production*, «Psychological Review», 93(3), 283–321.

FERREIRA V. S. (2010), *Language production*, «Wiley Interdisciplinary Reviews: Cognitive Science», 1, 834–844.

FODOR J.A. (1976), *The language of thought*, Harvester Press, Sussex, UK.

FODOR J.A. (1981), *The Present Status of the Innateness Controversy* in Id. (ed.), *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*, MIT Press, Cambridge, MA, pp. 257–316

FODOR J.A. (1998), *Concepts. Where cognitive science went wrong*, Oxford University Press, Oxford.

FODOR J.A., FODOR *J.D.,* GARRETT M. (1975), *The psychological unreality of semantic representations*, «Linguistic Inquiry», 6, 515–531.

FODOR J.A., GARRETT M., WALKER E., PARKES C., (1980) *Against definitions*, «Cognition», 8, 263–367.

GARRETT M.F, (1975). *Syntactic process in sentence production*, in BOWER G. (Ed.), *The Psychology of Learning and Motivation*, Vol. 9, Academic Press, New York, pp. 133–177.

HAMPTON J.A., (1982), *A demonstration of intransitivity in natural categories*, «Cognition», 12(2), pp. 151–164.

HART J.JR., BRENDT R.S., CARAMAZZA A., (1985), *Category–specific naming deficit following cerebral infarction*, «Nature», 316(6027):439–40.

HARVEY D.Y., SCHNUR T.T, 2016, *Different Loci of Semantic Interference in Picture Naming vs. Word–Picture Matching Tasks*, «Frontiers in Psychology», 7: 710.

HOWARD D., NICKELS, L., COLTHEART M., COLE–VIRTUE J. (*2006*), *Cumulative semantic inhibition in picture naming: Experimental and computational studies*, «*Cognition», 100, 464–482*.

JACKENDOFF R. (1983)*, Semantics and Cognition*, MIT Press, Cambridge MA.

JACKENDOFF R. (1987)*, Consciousness and the Computational Mind*, Bradford/MIT Press, Cambridge, MA.

JACKENDOFF R. (1990*), Semantic Structures*, MIT Press, Cambridge, MA.

INDEFREY P. (2011), *The spatial and temporal signatures of word production components: a critical update*, «Frontiers in Psychology», 12; 2:255.

INDEFREY P., LEVELT W.J.M., (2004), *The spatial and temporal signatures of word production components*, «Cognition», 92, 101–144.

LAMBON RALPH M.A., SAGE K., JONES R. W., MAYBERRY E. J. (2010), *Coherent concepts are computed in the anterior temporal lobes*, «Proceedings of the National Academy of Sciences»*, 107,* 2717–2722.

LEVELT W.J.M. (1989), *From Intention to Articulation*, MIT Press, Cambridge, MA.

LEVELT W.J.M., ROELOFS A., MEYER A.S., (1999), *A theory of lexical access in speech production*, «Behavioural and Brain Sciences», 22, 1–75.

MEYER A.S., DAMIAN M.F., (2007*)*, *Activation of distractor names in the picture–picture interference paradigm*, «Memory & Cognition», 35, 494–503.

MORSELLA E., MIOZZO M., (2002), *Evidence for a cascade model of lexical access in speech production*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 28, 555–563.

OPPENHEIM G.M., DELL G.S., SCHWARTZ M.F., (2010), *The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production*, «Cognition», 114(2):227–52.

PATTERSON K., NESTOR P. J., ROGERS T. T. (2007), *Where do you know what you know? The representation of semantic knowledge in the human brain*, «Nature Reviews Neuroscience»*, 8,* 976–987.

PERCONTI P.*,* (2001)*, I disturbi lessicali nelle afasie,* in PENNISI A., CAVALIERI R., *Patologie del linguaggio e scienze cognitive,* Il Mulino, Bologna, pp. 161–192.

PUTNAM H., (1975), *The meaning of 'meaning'*, «Minnesota Studies in the Philosophy of Science», 7,131–193.

ROELOFS A., (1992), *A spreading–activation theory of lemma retrieval in speaking*, «Cognition»*, 42*, 107–142.

ROELOFS A., (1997). *A case for non decomposition in conceptually driven word retrieval*, «Journal of Psycholinguistic Research»*, 26*, 33–67.

ROELOFS A., (2008), *Tracing attention and the activation flow in spoken word planning using eye movements*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», *34*, 353–368.

ROELOFS A., (2014), *A dorsal–pathway account of aphasic language production: The WEAVER++/ARC model*, «Cortex»*, 59*, 33–48.

ROELOFS A., (2018). *A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming*, «Cognition»*, 172*, 59–72.

ROELOFS A., MEYER A.S., LEVELT W.J.M. (1998), *A case for the lemma–lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997)*, «Cognition», *69*, 219–230.

ROELOFS A., FERREIRA V.S. (in press), *The architecture of speaking*. In P. HAGOORT (ed.). *Human language: From genes and brains to behavior*, MIT Press, Cambridge, MA.

ROSCH E., MERVIS C.B., GRAY W., JOHNSON D., BOYES–BRAEM P., (1976), *Basic Objects in Natural Categories*, «Cognitive Psychology», 8, 3, 382–439.

SMITH M., WHEELDON L. (2004), *Horizontal Information Flow in Spoken Sentence Production*, «Journal of Experimental Psychology: Learning, Memory, and Cognition», 30(3), 675–686.

STARREVELD P. A., LA HEIJ W. (2004)*, Phonological facilitation of grammatical gender retrieval*, «Language and Cognitive Processes», 19 (6): 677–711.

VAN TURENNOUT M., HAGOORT P., BROWN C.M., (1998)*, Brain activity during speaking: from syntax to phonology in 40 milliseconds*, «Science», 280(5363):572–4.

VELARDI A., (2009), *I problemi della transitività nel trattamento computazionale delle categorie semantiche e dell'accesso lessicale*, in FERRARI G., MOSCA M., BENATTI R., (a cura di), *Linguistica e modelli teconologici della ricerca*, Bulzoni, Roma, pp. 597–603.

VELARDI A., (2005), *Il nuovo paradigma. Categorie, prototipi e semantica cognitiva*, EDAS, Messina.

VIGLIOCCO G., ANTONINI T., GARRETT M. F., (1997), *Grammatical gender is on the tip of Italian tongues*, «Psychological Science», 8(4), 314–317.

Concetti e processi di categorizzazione
ISBN 978-88-255-1306-6
DOI 10.4399/978882551306614
pp. 299-301 (marzo 2018)

Per concludere e continuare

di Francesco Gagliardi[1],
Marco Cruciani[2], Andrea Velardi[3]

I contributi raccolti in questo volume rappresentano alcuni dei temi più interessanti e attuali relativi allo studio dei concetti e dei processi di categorizzazione della mente umana. Nel seguito indichiamo, senza pretesa di esaustività, alcune opere la cui lettura potrebbe giovare al lettore interessato ad avere un panorama più ampio delle analisi e delle conoscenze relative all'oggetto di questa raccolta.

Tra le opere di consultazione a carattere generale con una buona voce dedicata ai concetti che può costituire un ottimo punto di partenza segnaliamo Houde (1998), Kruschkea (2001), Medin e Aguilar (1999).

Tra le monografie segnaliamo i lavori di Murphy (2002), Fodor (1998) e Thagard (2005), la prima indaga la natura dei concetti da un punto di vista quasi esclusivamente psicologico, la seconda è una fondamentale opera filosofica, la terza si caratterizza per un approccio più interdisciplinare proprio delle scienze cognitive.

Tra gli altri lavori con un'impostazione interdisciplinare che riteniamo di utile lettura, segnaliamo Cohen e Lefebvre (2005), Cordeschi e Frixione (2011), Gagliardi (2009, 2014).

[1] Independent Scholar, ORCID: 0000–0002–4270-1636. E–mail: fnc.research@gmail.com
[2] Università di Trento. E–mail: marco.cruciani@unitn.it
[3] Università di Messina. E–mail: velardi.velardi@gmail.com

Riferimenti bibliografici

COHEN, H. LEFEBVRE, C. (2005) (eds.) *Handbook of Categorization in Cognitive Science*, Elsevier Science Ltd, Oxford.

CORDESCHI, R., FRIXIONE, M. (2011) *Rappresentare i concetti: filosofia, psicologia e modelli computazionali*, «Sistemi Intelligenti» 23(1):25–40.

FODOR, J. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford University Press, New York.

GAGLIARDI, F. (2009) *La categorizzazione tra psicologia cognitiva e machine learning: perché è necessario un approccio interdisciplinare,* «Sistemi Intelligenti», 21(3):489–501.

GAGLIARDI, F. (2014) *La naturalizzazione dei concetti: aspetti computazionali e cognitivi*, «Sistemi Intelligenti», 26(2):283–295.

HOUDE, O. (1998) *Categorization*, in Houde, O., Kayser, D., Koenig, O., Proust, J., Rastier, F. (eds.) *Vocabulaire de sciences cognitives*, Presses Universitaires de France, Paris. (Traduzione italiana: Houde, O. *et al*. (2000) *Dizionario di scienze cognitive. Neuroscienze, psicologia, intelligenza artificiale, linguistica, filosofia*, Editori Riuniti, Roma.

KRUSCHKEA, J.K. (2001) *Categorization and Similarity Models*, in Smelser, N.J., Baltes, P.B. (eds.) *International Encyclopedia Of The Social & Behavioral Sciences*, Pergamon Press, Oxford, UK, pp. 1532–1535.

MEDIN, D.L., AGUILAR, C. (1999) *Categorization*, in Wilson, R.A., Keil, F. (eds.) *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, MIT Press, Cambridge, MA, pp. 104–106.

MURPHY, G.L. (2002) *The big book of concepts, MIT Press, Cambridge*, MA.

THAGARD, P. (2005) *Mind: Introduction to cognitive science, 2nd edn.*, MIT Press, Cambridge, MA.

LA MENTE E I SISTEMI COGNITIVI

Collana di scienze cognitive, filosofia e tecnologia

1. Marco Cruciani
   Il ruolo della conoscenza fattuale nella determinazione del significato
   ISBN 978-88-255-0526-9, formato 14 × 21 cm, 124 pagine, 10 euro

2. Francesco Gagliardi, Marco Cruciani, Andrea Velardi
   Concetti e processi di categorizzazione
   ISBN 978-88-255-1306-6, formato 14 × 21 cm, 304 pagine, 16 euro

Compilato il 30 marzo 2018, ore 14:40
con il sistema tipografico LaTeX 2_ϵ

Finito di stampare nel mese di marzo del 2018
dalla tipografia «System Graphic S.r.l.»
00134 Roma – via di Torre Sant'Anastasia, 61
per conto della «Gioacchino Onorati editore S.r.l. – unipersonale» di Canterano (RM)