

MARC ARTIGA

Universitat de València

Es pot naturalitzar la intencionalitat?

Resum: La intencionalitat és un element central en la nostra comprensió del món. Diàriament atribuïm estats intencionals (com ara creences, desitjos o percepcions) per a poder explicar el comportament d'altres agents, i moltes teories filosòfiques hi apel·len per a entendre altres nocions més complexes. Tanmateix, els estats intencionals són entitats enigmàtiques. Aquest article explica què és la intencionalitat i per què és tan important i problemàtica alhora. En segon lloc, examina diverses teories naturalistes que intenten mostrar que la intencionalitat és compatible amb una visió científica del món. Finalment, com que totes les propostes existents s'enfronten a dificultats significatives, s'exploren les opcions disponibles en cas que cap dels projectes naturalistes pugui reeixir.

Abstract: Intentionality is a central feature of our understanding of the world. We daily attribute intentional states (like beliefs, desires or perceptual states) to explain the behavior of other agents, and many theories appeal to them to understand more complex notions. Nonetheless, intentional states are puzzling entities. This article explains what intentionality is and why it is so important and problematic at the same time. Secondly, it examines various naturalistic theories, which seek to show that intentionality is compatible with a scientific worldview. Finally, given that all extant proposals face significant difficulties, it explores the available options in case no naturalistic theory can succeed.

Paraules clau: representació, funció, explicació, biologia, ciència cognitiva.

Keywords: representation, function, explanation, biology, cognitive science.

INTRODUCCIÓ

SUPOSEM QUE UN DIA l'Empar va al cinema. Si volem explicar per què hi ha anat, podem apel·lar al fet que *creia* que hi havia una bona pel·lícula a la cartellera i que *desitjava* veure-la. Si, a més a més, volem explicar per què creia que hi havia una bona pel·lícula, potser mencionarem que va *veure* el títol a l'entrada del cinema i va *recordar* que un amic seu li havia *dit* que era molt bona. En aquestes explicacions, usem verbs com 'creure', 'desitjar', 'veure' o 'recordar', que tenen certes característiques en comú. En primer lloc, aquestes expressions es refereixen a estats mentals: les creences, les experiències, les sensacions o els records són estats de la ment (en contraposició amb l'alçada o la calvície, per exemple). Però també comparteixen una altra propietat: tots aquests estats mentals estan dirigits a alguna altra cosa. Quan creiem, desitgem o recordem, sempre creiem, desitgem, o recordem *quelcom*. El desig és sempre l'anhel d'alguna cosa, i per saber és necessari que hi hagi alguna cosa sabuda. Les creences, els records, el coneixement i les percepcions, així com molts altres estats mentals, necessiten estar relacionats amb alguna cosa. L'objectiu principal d'aquest assaig és procurar entendre una mica millor aquesta propietat tan important i, alhora, tan estranya de nosaltres mateixos.

De fet, si hi pensem un moment, ràpidament ens adonarem que no només els estats de la ment estan dirigits a un objecte. Per exemple, a l'anterior explicació del comportament de l'Empar també apel·làvem a allò que un amic li va *dir*. Les expressions lingüístiques (com totes les d'aquest text) són també entitats que refereixen a alguna altra cosa. L'expressió 'El riu Xúquer desemboca a Cullera' refereix a un fet del món, a saber, que el riu més llarg del País Valencià troba el mar a Cullera. L'oració serà vertadera o falsa depenent de si el riu efectivament desemboca en aquesta ciutat costera. I no només les expressions lingüístiques posseeixen direccionalitat. Pensem en les imatges; un famós quadre de Joaquim Sorolla, per exemple, representa uns nens estirats a la sorra de la platja. Per tant, en algun sentit, el quadre també està dirigit a un fet del món, versa sobre un esdeveniment. Això contrasta, per exemple, amb les pedres, els núvols o el vent que normalment no representen res i, per tant, no poden ser ni vertaders ni falsos.

Històricament s'han utilitzat moltes expressions per parlar d'aquesta propietat sorprenent: 'estar dirigit a', 'significar', 'referir', 'denotar' o 'representar' en són alguns exemples. En filosofia contemporània, l'expressió més general que engloba tots aquests usos i que deriva de Franz Brentano és el d'"intencionalitat". En ciència, l'ús més estès és el de 'representar'. Aquí usarem els dos termes indistintament. Per tant, direm que les expressions lingüístiques, les pintures o els estats mentals són representacions o posseeixen intencionalitat.

Són estats que representen alguna cosa i que típicament poden ser vertaders (correctes) o falsos (incorrectes) segons l'existència d'allò representat. Com veurem, aquesta relació és molt més important del que podríem pensar a primera vista i, tanmateix, planteja un conjunt de problemes de difícil resposta. Aquest article analitza què és la intencionalitat i si l'existència d'estats intencionals és compatible amb una visió científica del món.

I. LA INTENCIONALITAT: QUÈ ÉS I QUINS PROBLEMES PLANTEJA?

Els estats intencionals juguen un paper molt important en les nostres vides. En primer lloc, com l'anterior exemple de l'Empar pretenia mostrar, són essencials en les explicacions habituals del comportament humà (el que en filosofia s'anomena 'psicologia popular' o *folk psychology*). Si volem explicar per què el Miquel va caminar cap a la nevera, va obrir-la, va treure una cervesa i se la va beure, segurament apel·larem al fet que *desitjava* una cervesa i *creia* que n'hi havia una a la nevera. L'atribució d'estats intencionals a altres persones ens permet entendre i predir el seu comportament. Convido el lector a intentar explicar les accions de les persones que l'envolten sense adscriure'ls experiències, records, creences o desitjos. Segurament li serà extremadament difícil, si no impossible.

Ara bé, l'atribució d'estats intencionals no està restringida a la psicologia popular, sinó que també és una peça fonamental en diverses disciplines, com ara les ciències cognitives i la biologia. D'una banda, la psicologia científica explica moltes de les nostres capacitats cognitives (la percepció, el pensament, el domini del llenguatge, etc.) atribuint un gran nombre de representacions. Un cas paradigmàtic és la lingüística actual, hereva de la revolució chomskiana, que explica la nostra habilitat per a entendre, processar i emetre expressions lingüístiques atribuint-nos un gran nombre de representacions mentals sobre la gramàtica i el lèxic. El mateix es pot dir de la neurociència. Per exemple, segons la teoria més estesa en aquesta àrea, la nostra capacitat per reconèixer cares involucra tres regions fonamentals del còrtex cerebral: l'Àrea Facial Fusiforme (AFF), l'Àrea Facial Occipital (AFO) i el Solc Temporal Superior posterior (STSp). Disposem de suficient evidència per pensar que la AFO és responsable de la representació dels elements que típicament configuren una cara: nas, orelles, barbata... Aquesta informació passa a la AFF, que probablement és la part del cervell que s'ocupa del reconeixement facial. La AFF identifica cares a partir de detectar propietats més estructurals, com ara la forma, la relació entre els diferents components, etc. (HAXBY *et al.* 2002; FAIRHALL i ISHAI 2007; ISHAI 2008; KANWISHER i BARTON 2010). Per tant, un element essencial

en l'explicació de com els humans som capaços de reconèixer cares involucra estats cerebrals que representen certes propietats del món (nas, boca, configuració, etc.), i mecanismes que usen i transformen aquestes representacions. De fet, no és exagerat afirmar que la noció de representació és la pedra angular que connecta les diferents àrees que han passat a anomenar-se 'ciències cognitives' (STERNBERG i STERNBERG 2009).

També en biologia aquesta noció és cabdal. Pensem, per exemple, en la capacitat d'orientar-se en l'espai que tenen molts animals no humans. Les formigues del desert *Catalgyphis bicolor* coneixen la seva posició en un determinat moment a partir d'un càlcul que requereix, d'una banda, recordar la direcció en què s'han desplaçat des del punt d'origen (que detecten a partir de la posició del sol o la llum polaritzada) i, d'altra, registrar tots els passos que fan. És com si cada matí contéssim les passes que fem des que sortim de casa i ens fixéssim en la direcció en què caminem, i utilitzéssim aquesta informació per calcular la nostra ubicació actual. Aquest mecanisme usat per les *Catalgyphis bicolor* i molts altres animals, anomenat '*path integration*', òbviament requereix crear vàries representacions sobre la pròpia posició, els moviments de les cames, etc., i, a més a més, realitzar certes operacions matemàtiques. Sense dubte, l'atribució de representacions és indispensable per a l'explicació del seu comportament.

Un altre mecanisme d'orientació espacial utilitzat per alguns animals com ara les rates (el descobridor del qual va guanyar un premi Nobel) són les anomenades 'cèl·lules de lloc'. Les cèl·lules de lloc són un conjunt de neurones de l'hipocamp (estructura cerebral compartida pels humans i altres vertebrats) amb una activitat molt forta quan la rata és en un lloc determinat de l'espai. Per exemple, hi ha certes neurones que mostren una gran activació en el moment en què l'animal es troba a la cantonada nord-est de l'habitació, unes altres neurones quan es troba a la cantonada sud-est, unes altres quan és al centre... I, com que hi ha cèl·lules específiques pels diferents indrets de l'espai, les rates desenvolupen un mapa cognitiu de l'habitació en què es troben. De fet, recentment s'ha descobert que també hi ha cèl·lules que es desapareixen quan troben un límit (per exemple, una paret); d'altres que usen com a marc la posició del cap de l'animal o, fins i tot, seguint un patró de xarxa, com si davant d'un tauler d'escacs certes neurones només s'activessin a les caselles negres (BOCCARA *et al.* 2010; MOSER *et al.* 2008). De nou, en totes aquestes explicacions els biòlegs atribueixen un gran nombre de representacions cerebrals. Això no només és imprescindible per a entendre com certs organismes són capaços de realitzar tasques molt complexes, sinó que també és crucial per a guiar la investigació científica i establir noves preguntes i projectes de recerca (BECHTEL 2016).

Finalment, cal mencionar que la noció de representació no és només important perquè moltes explicacions científiques i no científiques l'usen. Un motiu addicional per a voler entendre millor aquest fenomen és que moltes teories recorren a les representacions per a explicar altres nocions (aparentment) més complexes. Per exemple, un projecte molt ambiciós que ha rebut molta atenció en els darrers vint anys és el d'intentar explicar la consciència apel·lant a la intencionalitat. Què és el dolor? Què és l'alegria? Què és la sensació de vermellor que sentim quan veiem una poma madura? Com pot ser que un cervell constituït bàsicament per neurones pugui produir tots aquests estats conscients? Una possible resposta afirma que tenir l'experiència de vermellor quan veiem una poma no és res més que representar que la poma té una determinada propietat (TYE 2000). Paral·lelament, la sensació de dolor no seria sinó una representació de dany (BAIN 2003) i sentir por es reduiria a representar que alguna cosa pot causar-te un mal (PRINZ 2004). Òbviament, una proposta que intenti reduir o explicar els estats conscients a partir dels estats representacionals necessita fonamentar-se en una bona teoria naturalista de la representació. Per tant, aquests projectes també requereixen tenir una idea clara del fenomen de la intencionalitat.

Malauradament, malgrat l'ús tan estès d'aquesta noció en contextos científics i filosòfics, les representacions són entitats desconcertants. Però, algú podria objectar, què hi ha de misteriós? Certament, en algun sentit, les representacions sempre estan dirigides a algun objecte, però —diria l'objector imaginari— això simplement vol dir que són entitats relacionals. És a dir, *representar* és un tipus de relació entre dues entitats: quan el Vicent imagina un gelat de mango i kiwi, s'estableix una relació (*imaginar*) entre la persona (el Vicent) i l'objecte (el gelat de mango i kiwi). Això és exactament el que passa amb altres relacions, com ara *ser al costat de* o *ser més alta que*. Si la Joana és al costat de la Laia, també hi ha una relació (*ser al costat de*) entre dues entitats (Joana i Laia); de la mateixa manera, si la Sagrada Família és més alta que el Miquelet, hi ha una relació (*ser més alta que*) entre aquests dos edificis. Tot això és ben cert. I, tanmateix, aquest paral·lisme aparent amaga algunes diferències substancials. N'assenyalarem breument tres de les que fan que aquesta sigui una noció extremadament problemàtica: la inexistència de l'objecte intencional, la normativitat i la incògnita dels fonaments.

En primer lloc, les relacions normalment pressuposen l'existència de les coses relacionades: si la Sagrada Família o el Miquelet no existissin, un no podria ser més alt que l'altre i el mateix es podria dir sobre la Joana i la Laia. En canvi, A pot representar B encara que B no existeixi. Per exemple, jo puc creure que el monstre de Banyoles té ales o que el fill de Wittgenstein és filòsof encara que ni el monstre ni el fill de Wittgenstein no hagin existit mai. I el Vicent es

pot imaginar un gelat de mango i kiwi, malgrat que mai n'hi hagi hagut cap. El fet que puguem representar entitats inexistents és una de les propietats més importants dels estats intencionals que els diferencien de molts altres estats, i que és molt difícil d'explicar. Suposant que mai ha existit cap gelat de mango i kiwi, com és que el Vicent pot representar tal cosa? Com podem estar relacionats amb coses que no existeixen?¹

En segon lloc, els estats intencionals típicament poden ser vertaders o falsos, correctes o incorrectes, etc. Per exemple, la creença que la terra és el centre absolut de l'univers és una creença falsa i, si tinc una il·lusió òptica (per exemple, la famosa il·lusió Müller-Lyer, figura 1), la meua percepció és incorrecta.² De manera semblant, l'expressió lingüística 'la Joana és al costat de la Laia' pot ser vertadera o falsa. En canvi, el fet que la Joana sigui al costat de la Laia no té valor de veritat; no pot ser avaluat com a vertader (correcte) o fals (incorrecte).³ El fet que un gerro reposi sobre una taula no és vertader ni fals, simplement és així; per contra, l'oració 'el gerro és sobre la taula', sí que té valor de veritat. Aquesta propietat normalment s'anomena la 'normativitat' dels estats intencionals, perquè estableix una norma a partir de la qual les representacions són avaluades. Però, per què les relacions *ser al costat de* o *més alt que* no poden ser avaluades normativament i, en canvi, les representacions sí? I, encara més difícil, quina és la naturalesa d'aquestes normes? És l'existència de normes compatible amb una visió científica del món?

¹ Una altra propietat relacionada amb l'objecte intencional és el que s'anomena intensionalitat (amb 's'). Si jo sóc al costat d'en Joan i en Joan és un agent secret del govern rus, llavors és veritat que jo sóc al costat d'un agent del govern rus (encara que no ho sàpiga). En canvi, del fet que jo cregui que en Joan és alt, i que ell sigui un agent secret del govern rus no es pot deduir que jo cregui que és un agent secret del govern rus. Una mica més tècnicament, la intensionalitat té a veure amb el fet que si canviem un element de la representació ('Joan') per un element amb la mateixa referència ('agent secret del govern rus'), pot ser que el valor de veritat de la representació canviï ('el Marc creu que el Joan és alt' és veritat, però 'el Marc creu que un agent secret rus és alt' és fals). Sobre la relació entre intensionalitat i la naturalització de la intencionalitat, vegeu Fodor (1990).

² És important destacar, però, que no tots els estats intencionals poden ser avaluats com a correctes o incorrectes. Per exemple, tot i què 'aquesta cadira és de fusta' pot ser vertader o fals, 'fusta' no té valor de veritat. Alguns estats mentals com 'imaginar' també semblen quedar al marge.

³ Evidentment, hom podria emprar altres normes (ètiques o pràctiques, per exemple) per a avaluar aquesta relació. Si hi ha alguna norma ètica que impedeixi que la Joana sigui al costat de la Laia, aquest fet serà *èticament* incorrecte. Però sembla que una creença o una oració, per exemple, és vertadera o falsa, correcta o incorrecta independentment de qualsevol altra valoració (ètica o pràctica) que fem. La creença en el geocentrisme és errònia, més enllà que en determinats moments històrics hi hagués raons pràctiques per a sostenir-la (recordeu Galileu!).

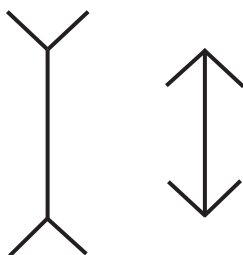


Fig 1: A la il·lusió Müller-Lyer, la línia central de la figura de l'esquerra sembla més llarga que la línia central de la figura de la dreta. Tanmateix, les dues tenen exactament la mateixa longitud.

Finalment, un altre tret distintiu dels estats intencionals és que és del tot misteriós què és el que fa que certs estats siguin representacions i d'altres no. Sembla obvi què fa que un objecte estigui al costat d'un altre o per què un objecte és més o menys alt, però per què les paraules representen alguna cosa? Per què les cadires normalment no són representacions i les ones sonores que emeto quan parlo, sí? Es podria suggerir que els sons que produeixo són representacions perquè els emeto amb la *intenció* de transmetre cert contingut. És a dir, hom podria intentar explicar la intencionalitat de les expressions lingüístiques apel·lant a les intencions, desitjos i creences dels agents que les emeten. De fet, aquesta és probablement la proposta més habitual (cfr. MILLIKAN 2005). I tanmateix, això només desplaça el problema perquè, com expliquem la intencionalitat dels estats mentals (que, al seu torn, usem per explicar la intencionalitat dels signes externs)? Per què un determinat estat cerebral representa quelcom? Què fa que un patró d'activació neuronal (i no un altre estat) representi la presència d'una cara (i no un altre objecte)? En aquest cas, no podem recórrer a un altre agent (un homenet intern?) que tingui la intenció o la creença de produir aquest estat cerebral, ja que evidentment això portaria a un regress a l'infinit. Com veiem, no és fàcil saber en virtut de quina propietat o procés certes entitats del món són representacions i d'altres no. Això és el que podem anomenar 'la incògnita dels fonaments'.

En resum, el fenomen de la intencionalitat recull un conjunt de característiques excepcionals. D'una banda, és un element essencial en les explicacions de la psicologia popular, la psicologia científica, la biologia, les ciències socials i la filosofia (entre d'altres) de tal manera que és extremadament difícil imaginar-se com podríem explicar el món que ens envolta sense adscriure-hi representacions. Tanmateix, són entitats aparentment misterioses. Podem representar entitats inexistents, les representacions estableixen normes de correcció

i incorrecció i, a més a més, no sabem en virtut de quin procés alguns estats posseeixen intencionalitat i d'altres no. Per tant, es tracta d'un problema filosòfic de primer ordre.⁴

Per a intentar resoldre aquestes enigmes, els filòsofs han presentat una sèrie de teories que normalment s'agrupen sota la denominació de 'teories naturalistes de la intencionalitat'. Com és ben sabut, l'expressió 'naturalisme' s'usa de maneres molt diverses, però en aquest context s'entén que una proposta naturalista pretén mostrar de quina manera els estats intencionals són compatibles amb la nostra visió científica del món. Més concretament, les teories naturalistes de la intencionalitat tenen com a objectiu explicar què és la intencionalitat apel·lant a entitats no misterioses i no intencionals (per evitar els problemes de regrés infinit que hem mencionat anteriorment). En definitiva, la idea és proporcionar una llista de condicions que determinin quan hi ha una representació i quin és el seu contingut (és a dir, què representa). Per il·lustrar el projecte, considerem breument el cas de verola. Durant molt de temps es va pensar que la verola era produïda per un geni maligne que s'introduïa dins del cos del malalt. Al segle xx, però, es va desenvolupar el que podríem anomenar una 'teoria naturalista de la verola': es va descobrir que, de fet, aquesta malaltia és causada per un virus del gènere *Orthopoxvirus*. Aquest descobriment va permetre explicar algunes de les característiques de la malaltia que es coneixien i en va poder desmentir d'altres, com la idea que el causant de la verola té por del color vermell i dels gossos. De manera similar, l'objectiu de les teories naturalistes de la intencionalitat és explicar què són els estats intencionals, i així vindicar algunes de les nostres idees sobre aquests estats i, de ben segur, desmentir-ne d'altres.

En aquest punt, però, és important fer una distinció (que, malauradament, molt sovint és obviada pels estudis especialitzats). Quan hom presenta una teoria naturalista, hi ha dues qüestions diferents que podria estar abordant. En primer lloc, hom podria preguntar-se per què una activació neuronal o una fotografia són representacions. Què diferencia una representació (per exemple, un estat mental) d'un estat que no es una representació (per exemple, una taula)? Anomenarem aquest problema la qüestió *metasemàntica*. Aquesta és la pregunta més fonamental, però també la més difícil i, com veurem, molts autors han preferit deixar-la en un segon terme a causa de la seva complexitat. En comptes d'això, han intentat oferir el que podem denominar una teoria *semàntica*: assumint que R és una representació, què fa que representi A o B? És a dir, pressuposant que una fotografia, per exemple, és una representació, què

⁴ Per a una explicació amb més detall dels estats intencionals i els problemes que plantegen, vegeu Pineda (2012, 36-57) i Jacob (2014).

fa que sigui una fotografia del Vicent, o una fotografia del Ximo? La qüestió metasetemàntica té a veure amb l'estatus representacional, mentre que la qüestió semàntica es pregunta pel seu contingut. Per a qualsevol representació R, un sempre es pot plantejar les dues qüestions: per què R representa A en comptes de B (qüestió semàntica) i per què R representa alguna cosa, en comptes de no representar res (qüestió metasetemàntica).⁵

A continuació analitzarem les teories més conegudes que intenten abordar la qüestió semàntica: les teories de la semblança, les teories causals i les teories informacionals. Després ens centrarem en les teories teleològiques, les úniques que ofereixen una resposta a la qüestió semàntica i metasetemàntica alhora. Veurem que cada una d'aquestes propostes té bones motivacions, però s'enfronta també a dificultats considerables. Acabarem amb una breu discussió de les opcions restants: l'eliminativisme, el ficcionalisme i el pluralisme.

2. TEORIES NATURALISTES DE LA INTENCIONALITAT

2.1 Teories de la semblança

Reflexionem un moment sobre les representacions icòniques, com ara les imatges. Per què, per exemple, els quadres de Canaletto representen Venècia? Per què la imatge del televisor és una representació d'un partit d'handbol? Per què el clauer que comprem a París amb la forma de la Torre Eiffel representa aquest edifici? Una resposta força intuïtiva és que són representacions d'aquests fets perquè s'hi assemblen. És a dir, el quadre de Canaletto representa els canals de Venècia perquè, en algun sentit, són similars. Seguint aquest raonament, es podria suggerir que allò que fa que una representació sigui sobre S i no sobre S* és que s'assembla a l'una i no a l'altra. De fet, aquesta idea és molt antiga: una de les figures més destacades que va intentar explotar-la per explicar la intencionalitat dels estats mentals és John Locke (1690/1990), tot i que idees similars podrien trobar-se ja en el *De anima* d'Aristòtil (SHIELDS 2016). Seguint aquesta intuïció, una formulació general de la *teoria de la semblança* (TS) seria la següent:

(TS) Un estat R representa S si i només si:

- (1) R és una representació
- (2) R s'assembla a S

⁵ Tot i que la distinció conceptual entre els dos projectes és clara, com veurem més endavant no hauríem de pressuposar acríticament que és possible donar una teoria satisfactòria d'una d'aquestes qüestions sense donar una bona resposta a l'altra.

Abans d'entrar a valorar aquesta proposta, cal tenir en compte que, com hem dit abans, TS és una teoria semàntica (no metasemàntica) de la representació. TS assumeix que certs estats són representacions (tal com diu la condició 1) i intenta explicar com es determina el seu contingut. Això té tot el sentit del món, perquè seria una temeritat provar de definir què és una representació apel·lant a meres relacions de similitud: la immensa majoria d'entitats del món s'assemblen a alguna cosa, però només algunes són representacions. Una fulla de pi se n'assembla a una altra, però no la representa. Per tant, TS aborda exclusivament la qüestió semàntica.⁶

Tanmateix, perquè TS sigui informativa, caldria especificar amb més detall quina és la relació de semblança mencionada a la condició 2. Què fa que dues entitats siguin similars? D'altra banda, assumint que les entitats s'assemblen i es diferencien en diversos aspectes, quines relacions de semblança són les rellevants? En alguns casos, la resposta a aquestes preguntes pot semblar evident. El clauer de la Torre Eiffel s'assembla a l'original perquè hi comparteix el color, la textura i la forma, tot i que té una mida radicalment diferent. Tanmateix, en molts altres casos la resposta és lluny de ser òbvia. No és clar, per exemple, de quina manera la percepció d'un gat (l'estat mental causat per la presència d'aquest animal quan se'n és presentat a plena llum del dia i en plenes facultats) s'assembla al gat. En què s'assembla el patró d'activació neuronal a l'objecte percebut? De fet, fins i tot si pensem en els casos paradigmàtics que motiven aquesta teoria ens adonarem que la noció de semblança és menys transparent del que podríem suposar. Un nen pot assemblar-se a un altre nen (els dos tenen cap, braços, nas, etc.), però en quin sentit les taques de pintura sobre el llenç de Sorolla s'assemblen a un nen? Què té en comú un quadre tacat de pintura de colors i un nen a la platja? A la pintura, no hi ha nassos, ni sorra, ni aigua; a la platja, no hi ha pintura de colors. És difícil veure com una definició general de semblança pot explicar aquest fet.⁷

Una proposta interessant que s'ha suggerit sovint per a intentar resoldre aquestes qüestions és buscar una semblança de tipus estructural. Tot i que una pintura d'un nen i un nen no comparteixen ni textura, ni mida, ni cap altra propietat fàcilment identificable, hi ha una certa similitud estructural entre la

⁶ Aquest raonament pot semblar trivial, però ofereix una resposta a una de les objeccions més comunes a la teoria de la semblança. És habitual objectar a TS que si R s'assembla a S, llavors S s'assembla a R, mentre que en general no és cert que si R representa S, S també representa R. La raó per la qual aquest argument no funciona és que TS *assumeix* que R és una representació i S no. Per tant, tot i que S s'assembla a R, TS no implica que S l'ha de representar.

⁷ Evidentment, quan nosaltres mirem la pintura, percebem una semblança entre ella i la realitat, però de nou, el naturalista no pot apel·lar als nostres estats intencionals (percepcions, creences, etc.) per definir què són els estats intencionals.

representació i la cosa representada. Per exemple, de la mateixa manera que la cara del nen té boca, nas i ulls, a la pintura hi ha una taca diferent per cada un d'aquests elements. Extrapolant aquesta idea, l'estratègia és identificar una estructura compartida entre la representació i allò representat, de tal manera que entre la pintura i el nen hi hagi un isomorfisme.⁸ Aquesta és una idea molt poderosa que han explotat alguns filòsofs de les ciències cognitives (entre d'altres, CUMMINS 1996; CHURCHLAND 2001; RAMSEY 2007). Per exemple, sembla raonable pensar que les cèl·lules de lloc que les rates usen per a orientar-se tenen una relació estructural amb l'espai representat: cada neurona es correspon a un indret de l'habitació, de manera que les relacions entre neurones són isomòrfiques a les relacions que hi ha entre els espais representats. Aquesta és la raó per la qual un mapa cognitiu és tan útil. Malauradament, però aquesta proposta té el problema invers a la versió anterior: si bé és molt difícil trobar alguna propietat no estructural comuna entre les taques de pintura i un xiquet, és excessivament fàcil trobar relacions estructurals entre el quadre i qualsevol objecte. Permeteu-me explicar-ho amb més detall.

En general, un isomorfisme es defineix com una correspondència 1-1 entre dos conjunts. Si tinc dos conjunts amb tres objectes cada un, $A = \{1,2,3\}$ i $B = \{a,b,c\}$, la funció F que relaciona els objectes de la manera següent és un isomorfisme: $\{ \langle 1,a \rangle, \langle 2,b \rangle, \langle 3,c \rangle \}$. Tanmateix, les relacions següents també són isomorfismes: $\{ \langle 1,b \rangle, \langle 2,c \rangle, \langle 3,a \rangle \}$, $\{ \langle 1,c \rangle, \langle 2,b \rangle, \langle 3,a \rangle \}$. Entre dos conjunts qualsevol que tinguin el mateix nombre d'objectes, sempre és possible traçar diferents isomorfismes;⁹ i, si dos conjunts tenen el mateix nombre d'objectes, segur que hi ha un isomorfisme entre ells. Per exemple, hi ha molts isomorfismes entre les cares d'un dau i les potes d'un insecte (tots dos conjunts tenen 6 elements), o entre una dotzena d'ous i el mesos de l'any. Per aquesta raó, és d'esperar que hi hagi un isomorfisme entre la pintura de Sorolla i molts objectes —no només els nens que eren a la platja—. En conseqüència, una teoria que interpreti la condició 2 de TS com a semblança estructural (i l'analitzi simplement com a isomorfisme) és extremadament liberal, ja que implicaria que tota representació té com a contingut un nombre excessiu d'entitats. Evidentment, hom podria intentar reduir aquest nombre restringint el tipus d'isomorfisme, però això és lluny de ser una tasca fàcil (SHEA 2013A).¹⁰

⁸ Etimològicament, 'isomorfisme' vol dir 'igualtat de forma', del grec 'isos' (igual) i 'morphé' (forma, estructura).

⁹ Excepte el conjunt buit i el conjunt unitari, evidentment.

¹⁰ De fet, en lògica i matemàtiques sovint s'usa 'isomorfisme' per a referir-se a un relació que a més de ser 1-1 també preserva l'estructura de cada conjunt. Si aquesta relació més exigent és capaç de donar una resposta satisfactòria al problema del liberalisme és una qüestió oberta.

Per mor de l'argument, però, suposem que hi ha una noció de semblança que permeti limitar aquest liberalisme exacerbant. Tot i així, hi ha raons per a pensar que la relació de similitud entre representació i representat no és ni necessària ni suficient. Per exemple, imaginem que fotografio al Vicent, i que té un germà bessó idèntic que es diu Ximo. Certament, la fotografia representa al Vicent, però s'assembla tant a ell com al Ximo. Per tant, que hi hagi una relació de semblança (estructural o no estructural) entre A i B no és suficient perquè hi hagi representació de B (si fos així, la meua fotografia també representaria al Ximo). De la mateixa manera, imaginem una foto molt llunyana del Vicent, tan llunyana que gairebé només es veu com un punt negre enmig de la imatge. Diríem que aquest punt representa al Vicent però no s'hi assembla gens. Per tant, la semblança entre R i S tampoc sembla necessària perquè hi hagi una relació representacional.

De fet, aquests exemples de fotografies suggereixen una perspectiva diferent per resoldre el problema de la intencionalitat. En aquests casos, sembla que la raó per la qual la fotografia representa el Vicent (i no, per exemple, el seu germà bessó) és que jo vaig fer la foto quan el Vicent (i no el Ximo) estava davant. És a dir, sembla que hi ha una relació causal entre Vicent i la fotografia, que no existeix entre ella i el Ximo. De la mateixa manera, semblaria que un punt negre de la foto representa Vicent perquè és ell qui l'ha causada (independentment del grau de semblança). Si l'hagués causat el Ximo o un hipopòtam, el punt els representaria a ells. Aquest raonament ens porta directament a la teoria causal.

2.2 Teories causals

Hem vist com l'exemple de la fotografia suggereix que el contingut d'una representació podria venir determinat per una relació causal, però no és difícil trobar altres motivacions per aquesta proposta. D'una banda, els filòsofs amb certes inclinacions naturalistes usualment han recorregut a les relacions causals per a explicar moltes nocions difícils, com per exemple la teoria causal de la referència de Kripke (1980) o la teoria causal de la percepció (GRICE 1961). De l'altra, intuïtivament sembla que molt sovint els nostres estats mentals són causats precisament per les coses que representen; no hi ha dubte que la meua representació sensorial de color vermell freqüentment és produïda per coses d'aquest color. Per tant, algú podria estar temptat a defensar la teoria causal següent (TC):

(TC) R representa S si i només si:

- (1) R és una representació
- (2) S causa R

De nou, TC és una teoria semàntica; és a dir, no intenta explicar què són les representacions sinó per què un estat representacional té S i no S* com a contingut. Desgraciadament, però, TC té greus problemes que expliquen per què difícilment algú podria sostenir-la en aquesta forma nua. En veurem dos de fonamentals: el problema de l'error i el problema de la indeterminació.

Com dèiem al principi, una de les propietats definitòries dels estats intencionals és que poden ser falsos. R pot representar S quan S no s'esdevé. Per exemple, la creença 'les balenes són mamífers' o la il·lusió Müller-Lyer són representacions falses (o incorrectes). El fet que R pugui representar un estat inexistent és una de les propietats que distingeix els estats intencionals de molts d'altres. I, tanmateix, no està clar que TC sigui compatible amb aquesta característica. Què hauria de passar d'acord amb TC perquè R representi erròniament que S és el cas? S'haurien de complir les tres condicions següents: (1) R és una representació; (2) S causa R; i (3) S no és el cas. Les dues primeres clàusules s'haurien de complir perquè R representés S (segons TC) i la tercera és la condició que faria R falsa. Si els tres requisits es poguessin satisfer alhora, llavors TC seria compatible amb casos d'error. Malauradament, les condicions 2 i 3 no poden ser vertaderes al mateix temps, per la senzilla raó que només entitats que existeixen en l'espai i el temps poden ser causes (RODRÍGUEZ-PEREYRA 2014). Si una cosa no existeix, com podria ser la causa d'alguna altra? Per il·lustrar el problema, imaginem que tinc la percepció visual d'una pilota davant meu. Segons TC, la meua percepció pot representar aquest objecte només si de fet la pilota l'ha causada. Però, en aquesta cas, per força hi ha d'haver una pilota. Per tant, si R representa S, R ha de ser vertadera. Alternativament, si no hi hagués cap pilota davant meu, no podria causar el meu estat mental i, per tant, la meua percepció no podria representar-la. En conclusió, les condicions 2 i 3 no es poden satisfer mai alhora i, en conseqüència, si TC és veritat, les representacions mai no poden ser falses. Això és el que normalment s'anomena 'el problema de l'error'.

Una segona objecció (que de vegades s'ha confós amb la primera) és que el contingut que atribueix TC difereix significativament del contingut que intuïtivament haurien de tenir molts estats intencionals. Segons TC, un estat R representarà tota entitat que la causi. Tanmateix, és obvi que tot tipus de coses causen les nostres representacions: tot i que és possible que normalment els gossos causin el meu concepte gos,¹¹ TC requereix que *només* ho facin els

¹¹ Els conceptes són les representacions mentals que constitueixen els pensaments. Per exemple, quan penso que les granotes de l'estany són sorolloses, utilitzo els conceptes GRANOTA, ESTANY i SOROLLÓS.

Sobre la grafia, en aquest assaig seguiré la convenció habitual: quan parlem de conceptes usarem versaleta i quan parlem de continguts usarem cursiva. Per exemple, el contingut del concepte QUADRAT és *quadrat*. Tot i així, de vegades també usaré cursiva per introduir nocions o emfatitzar algun aspecte (espero que el context serveixi per evitar malentesos).

gossos. En concret, si al capvespre veig una guineu a certa distància i penso que és un gos, llavors una guineu haurà causat el meu concepte, i, per tant, hauríem de dir que el seu contingut no és només *gos*, sinó *gos i guineu*. Qualsevol cosa que causi el meu concepte automàticament formarà part del seu contingut representacional. D'aquesta manera, el meu concepte, que originàriament pensàvem que només es referia a gossos, de fet, denotaria una classe molt més àmplia d'entitats: *gos i guineu i gat gran i dibuix i...* Aquest és un contingut excessivament indeterminat. Sense dubte, les atribucions que fem habitualment en psicologia popular o en ciència tenen continguts molt més determinats (per exemple, el meu concepte GOS fa referència a gossos i no a guineus, gats grans, etc.). Per això, aquesta dificultat s'anomena normalment el 'problema de la indeterminació'.

Diverses teories causals han intentat esquivar aquestes dificultats, però fins ara cap a sigut capaç d'oferir una resposta plenament satisfactòria. D'entre les diverses propostes, sense dubte la més coneguda és la teoria de la dependència asimètrica, defensada per Fodor (1990). Per entendre-la, cal primer explicar la noció de 'dependència asimètrica'. Suposem que jo normalment saludo al Miquel quan me'l trobo pel carrer, però (com que sóc una mica miop) un dia veig un noi que s'hi assembla molt i el saludo també. En aquest cas diríem que la meua salutació al noi desconegut és *asimètricament dependent* de la salutació al Miquel en el sentit següent: si jo normalment no saludés el Miquel, no hauria dit adéu al noi desconegut, però si no saludés el noi desconegut, tot i així seguiria saludant el Miquel. És a dir, sembla que saludo al noi *perquè* sempre saludo el Miquel, mentre que no és veritat que saludo el Miquel perquè saludo aquest noi. Doncs bé, Fodor usa la noció de dependència simètrica per desenvolupar una teoria de la intencionalitat que solucioni els dos problemes anteriors: suposem que dues entitats, S i S*, causen R. Segons la teoria de la dependència asimètrica, R representa S (i no S*) quan la relació entre S* i R és asimètricament dependent de la relació entre S i R, és a dir, quan els condicionals següents són certs: si S* no causés R, S seguiria causant R, però si S no causés R, S* ja no causaria R. De manera planera, podem dir que R representa S i no S* quan trencar la connexió entre S i R trenca la relació entre S* i R però no al revés. Si pensem en l'exemple anterior, Fodor assenyalaria que tot i que les guineus de vegades causin el meu concepte GOS, el concepte refereix a gossos perquè la relació entre les guineus i el concepte GOS és asimètricament dependent de la relació entre els gossos i GOS: si els gossos no causessin GOS les guineus no ho farien, però si les guineus no causessin GOS, els gossos seguirien fent-ho.

Aquesta proposta sembla capaç de solucionar les dues dificultats anteriors. D'una banda, és compatible amb l'error perquè de vegades GOS pot ser causat per guineus tot i que no les representi. D'altra banda, també sembla

reduir la indeterminació del contingut: GOS no representa qualsevol causa, sinó només aquelles amb les quals té una relació que no depengui asimètricament de cap altra. No obstant això, no manquen contraexemples i objeccions (ADAMS i AIZAWA 1992, 2010; LOWER i REY 1993). Potser la més important és que Fodor mai ha estat capaç d'explicar per què certes relacions causals depenen d'altres, i l'explicació més intuïtiva fa caure la teoria en un cercle viciós. Per a entendre el problema, fem-nos la següent pregunta: per què la relació causal entre guineus i el concepte GOS depèn causalment de la relació entre gossos i GOS? Intuïtivament, sembla que això és així perquè GOS representa *gos* i no *guineu*. La relació asimètrica certament hi és, però s'explica pel fet que GOS té un determinat contingut. Conseqüentment, aquesta teoria pretén *explicar* el contingut representacional assumint precisament que les representacions ja tenen un determinat contingut i, com a resultat, cau en un cercle viciós.

Evidentment, hi ha altres propostes basades fonamentalment en relacions causals. Prinz (2000), per exemple, uneix una teoria de la dependència asimètrica amb la idea que la primera causa és la que determina el contingut (per a una crítica, vegeu ARTIGA 2014). Tanmateix, hi ha un altra aproximació al problema de la intencionalitat que també incorpora aspectes causals i que sembla capaç d'abordar els dos problemes fonamentals de TC: les teories informacionals.

2.3 Teories informacionals

Recentment, les propostes informacionals han guanyat molta popularitat tant entre filòsofs (SKYRMS 2010; BIRCH 2013; RUPERT 1999; SCARANTINO 2015) com també entre alguns científics (ELIASMITH 2000; USHER 2001). Hi ha dues motivacions principals per apel·lar a la noció d'informació. En primer lloc, és un terme àmpliament usat en les ciències cognitives —tot i que no sempre s'usi en el mateix sentit (FLORIDI 2010)—. A més a més, a diferència d'altres relacions més controvertides com la semblança o la causalitat, la informació que un estat porta sobre un altre es pot mesurar de manera molt precisa. De fet, hi ha una disciplina científica que s'ocupa d'aquesta tasca: la teoria de la informació. Això no només és important per satisfer l'objectiu de les teories naturalistes (que, recordem, és explicar la intencionalitat a partir d'altres entitats que coneguem millor), sinó que també ens permet establir fàcilment connexions entre aquestes propostes i les pràctiques científiques.

Però, abans de res, què és la informació? En quines condicions podem dir que un estat porta informació d'un altre? És habitual distingir dues maneres diferents d'entendre aquesta expressió. En primer lloc, quan parlem de la informació que adquirim quan llegim un diari o quan diem que tenim informació

confidencial, usem una noció *semàntica*¹² d'informació. En aquest sentit, un estat que porta informació no és sinó un estat intencional: aquesta informació pot ser vertadera o falsa (els diaris també s'equivoquen) i podem tenir informació sobre alguna cosa inexistent. Per tant, com que el nostre objectiu és precisament explicar què és la intencionalitat, difícilment la noció semàntica d'informació ens ajudarà a avançar en aquest projecte.

Tanmateix, sovint 'informació' no s'utilitza en un sentit semàntic, sinó en un de purament *quantitatiu*. Aquest és l'ús més habitual en teoria de la informació, on la informació normalment es mesura en bits. Intuïtivament, es pot dir que la informació que porta un estat depèn de quan redueix la incertesa d'un esdeveniment. Si un estat és molt improbable, el fet que s'esdevingui porta molta informació, mentre que un estat molt probable n'aporta menys. Per exemple, si aprenc que una moneda no esbiaixada llençada a l'aire ha caigut del costat de la cara, la meua incertesa s'ha reduït a la meitat (hi havia dues possibilitats, cara o creu, i ara només n'hi ha una) i, per tant, obtinc molta d'informació. En canvi, si la moneda estava molt esbiaixada cap a la cara, saber que n'ha sortit una me n'aportaria molta menys. A l'extrem, si la probabilitat que sortís cara fos màxima (per exemple, si la moneda té una cara a cada banda), llavors rebo 0 bits d'informació. Noteu que aquest tipus d'informació no pot ser ni vertadera ni falsa. La informació que aporta un estat es defineix per com de fet canvia les probabilitats prèvies. A més, la informació d'un estat es pot mesurar matemàticament. Això la converteix en un concepte molt atractiu tan per al científic com per al filòsof naturalista.

La diferència entre les dues nocions d'informació es pot entendre molt fàcilment amb l'exemple següent. Suposem que demà pot ploure o fer sol (però no les dues coses). Si el meteoròleg A ens diu que plourà i la meteoròloga B ens diu que farà sol, els dos ens aporten diferent *informació semàntica*, però la mateixa *quantitat d'informació* (redueixen la nostra incertesa a la meitat).

Ara bé, com podria usar-se la noció d'informació (quantitativa) per desenvolupar una teoria naturalista de la intencionalitat? De nou, la ciència ens suggereix una resposta, ja que en certes àrees de les ciències cognitives és habitual usar la informació d'un estat per establir quin és el seu contingut representacional. Un cas paradigmàtic són els estudis de Hubel i Wiesel realitzats als anys 50 i 60 en el còrtex visual de gats, i que han establert la metodologia emprada per tot una branca de la neurociència (HUBEL i WIESEL 1959). Aquests científics van introduir uns elèctrodes molt fins i precisos a través del crani

¹² Evidentment, la distinció que fem aquí entre una interpretació semàntica i una de quantitativa de la noció d'informació no té res a veure amb la distinció entre teories semàntiques i metasemàntiques que hem fet al principi.

de gats anestesiats¹³ per mesurar l'activitat elèctrica de neurones individuals, i d'aquesta manera van analitzar els patrons d'activitat d'una part significativa del còrtex visual, que és la part del cervell dedicada a processar la informació visual. Un cop introduït l'elèctrode, presentaven diversos estímuls davant dels ulls (cartons de diferents colors, mides, formes, etc.) i mesuraven l'activació d'una neurona determinada. Seguint aquest mètode, van descobrir que cada neurona era sensible només a una part del camp visual, és a dir, que només es disparava quan apareixia un estímulo en una àrea determinada del camp visual (per exemple, al voltant dels 30° a la part superior dreta). A més a més, van observar que algunes neurones tenen diferents patrons d'activació: n'hi ha que presenten una forta activitat quan hi ha un punt de foscrot envoltat de llum, i n'hi ha d'altres que presenten un alt grau d'activitat quan hi ha un punt de llum envoltat de foscrot. Hubel i Wiesel van anomenar aquests dos tipus de neurones 'on-center' i 'off-center'. Des de llavors, s'ha seguit el mateix mètode per a establir els camps receptius i els estímuls preferents de moltes neurones, tant al còrtex visual com fora d'ell (PURVES *et al.* 2008). De manera més general (i, evidentment, simplificada), aquesta metodologia ha estès la idea que si un conjunt de neurones R s'activa preferentment quan se li presenta un estímulo S, llavors R representa S. El contingut d'una representació es determina per l'estat que millor correlaciona amb ella. Ara bé, que R correlacioni amb S significa que quan R s'esdevé, es redueix la incertesa respecte a S i, per tant, R porta informació sobre S. En conseqüència, els científics que segueixen aquesta metodologia usen les relacions informacionals per determinar el contingut de les representacions.

Basant-se en aquest model, i generalitzant-lo a altres estats intencionals, alguns autors han defensat la següent *teoria informacional* (ELIASMITH 2000, 2005; USHER 2001):

- (TI) R representa S si i només si
- (1) R és una representació
 - (2) S és l'estat que més incrementa la probabilitat d'R
 - (3) R és la representació que més incrementa la probabilitat d'S¹⁴

Anem a pams. En primer lloc, la condició 1 simplement revela que TI és una teoria semàntica i no metasemàntica: assumeix que R és una representació i intenta establir-ne el contingut. Les condicions que realment fan la feina són

¹³ Aprofito per assenyalar que l'ús d'animals en l'experimentació científica planteja tot un seguit de debats ètics que malauradament no podem abordar aquí.

¹⁴ És a dir, de manera més precisa, R_i representa S_j si i només si (1) R_i és una representació i, per tot $j \neq i$, (2) $P(R_i/S_j) > P(R_i/S_i)$ i (3) $P(S_j/R_i) > P(S_i/R_i)$.

les 2 i 3. La clàusula 2 diu que la probabilitat d'R és més alta quan S s'esdevé que quan qualsevol altre estat és el cas i la clàusula 3 que la probabilitat d'S és més alta quan R s'esdevé que quan qualsevol altra representació és el cas. En altres paraules, entre tots els estats del món, S és el que millor prediu R, i entre totes les representacions, R és la que millor pronostica S. A tall d'exemple, el meu concepte GOS representa els gossos perquè (1) aquest concepte és una representació, (2) la presència d'un gos (i no la d'un gat o una guineu) és l'estat que més incrementa la probabilitat de que jo usi el concepte GOS i (3) l'ús del concepte GOS (i no el del concepte GAT o GUINEU) és la representació que més incrementa la probabilitat dels gossos. En el fons, les condicions 2 i 3 són una manera de capturar aquells estats que tenen una millor correlació amb la representació mental.

Com veurem tot seguit, TI sembla solucionar els problemes que afecten les teories causals, i segurament per això aquest tipus de propostes està guanyant molta popularitat. Tanmateix, mostraré que també s'han d'enfrontar a nous reptes.¹⁵

En primer lloc, TI soluciona el problema de l'error que hem presentat abans, ja que TI és clarament compatible amb el fet que algunes representacions siguin falses. La raó és molt senzilla: R pot ser l'estat que més incrementa la probabilitat de S i viceversa i, tot i així, pot passar que en una ocasió concreta R sigui el cas i S no. No obstant això, hi ha certs tipus d'errors que intuïtivament són possibles, però que contradiuen TI. Per exemple, quan les gallines veuen una ombra de certa mida desplaçant-se per terra, fugen corrents a amagar-se; l'explicació d'aquest comportament és que en certes ocasions l'ombra és causada per un falcó o algun altre depredador alat. Evidentment, la gran majoria de vegades fugen per no res (l'ombra la pot causar un núvol, algun ocell inofensiu, etc.). Suposem que la major part de les vegades la gallina fuig quan és sobrevolada per una inofensiva gavina, i només de tant en tant la fugida és causada per un falcó. A més, suposem també que les gavines sempre fan una ombra clarament visible, mentre que els falcons han desenvolupat certes estratègies per a evitar produir ombres recognoscibles. En aquest cas, la presència de gavines és el que més incrementa la probabilitat de l'estat mental d'alerta de les gallines, és a dir, la presència d'una gavina és l'estat que juga el rol d'S en la condició 2 de TI i, per tant, TI prediu que l'estat mental de les gallines quan veu una ombra representa que hi ha una inofensiva gavina. Això sembla una predicció

¹⁵ S'ha de subratllar, però, que queden alguns aspectes importants de la teoria que encara s'han d'aclarir. Per exemple, segons TI perquè un estat R en representi un altre, cal satisfer dues condicions. Però què passaria si es complís la clàusula 2 i no la 3, per exemple? Té la representació un contingut? Quin? Esperem que nous desenvolupaments d'aquestes teories ens permetin resoldre aquests i altres dubtes.

errònia; intuïtivament (i seguint la descripció dels etòlegs) aquest estat d'alerta representa la presència d'un falcó. De fet, si representés la presència de gavines, no podríem explicar per què les gallines fugen. Per què haurien d'escapar-se corrents, si les gavines són perfectament inofensives? Aquest problema òbviament apunta a una qüestió més general: basar el contingut de les representacions en correlacions és arriscat, perquè no és difícil imaginar certs contextos en què les representacions poden portar més informació sobre les entitats que no són representades (per exemple, si les gavines són comunes i no s'amaguen).

Pel què fa al problema de la indeterminació, novament TI sembla solucionar-lo: com que el referent d'una representació és l'estat que millor hi correlaciona, sempre hi haurà un únic estat representat. Per exemple, tot i que el meu concepte GOS de vegades és causat per guineus, segons TI aquest estat només pot representar una d'aquestes coses,¹⁶ i com que el concepte GOS segurament correlaciona més amb cans que amb guineus, és podria pensar que la teoria de la informació obté el resultat correcte. Tanmateix, aquesta solució també genera certes dificultats. Per mostrar-ho, suposem que en Joan reconeix la presència de gossos de la manera següent: quan percep una entitat que borda, llavors pensa que és un gos i activa el seu concepte GOS. Aquest mètode no és infal·lible, per què de vegades sent bordar en l'absència de cans, però funciona prou bé en el dia a dia. Ara bé, en aquest cas, resultaria que el concepte GOS correlacionaria millor amb la presència d'una cosa que borda que amb la presència de gossos. Això és així perquè els gossos de vegades resten muts, i per tant la probabilitat que en Joan pensi en gossos atès que hi ha un ca és més petita que la probabilitat que en Joan pensi en gossos si hi ha una cosa que borda. Encara pitjor: com que en Joan detecta la bordadissa a través de les ones sonores que estimulen els seus timpans, resulta que el concepte GOS correlaciona millor amb les ones sonores que estimulen els seus timpans que amb la presència d'una cosa que borda! La raó és que de vegades hi ha coses que borden que en Joan no sent, però si les sent segur que les ones estimulen els seus timpans.¹⁷ Per tant, TI implica que el concepte GOS de fet representa *ones sonores de tal i tal freqüència estimulant receptors auditius*. Evidentment, això és un resultat del tot

¹⁶ Cal remarcar que això pot ser la font d'un altre problema: la teoria informacional segurament és incompatible amb el fet que algunes representacions tinguin com a contingut diverses entitats (per la condició 2 de TI) i també amb el fet que diferents representacions tinguin el mateix contingut. Tanmateix és fàcil pensar en situacions dels dos tipus: un exemple del primer cas és el concepte BANC, que pot referir a una l'entitat financera o al seient allargat que pobla els parcs; del segon, podem pensar en els conceptes OCULISTA i OFTALMÒLEG.

¹⁷ Aquest raonament deriva d'un principi de la teoria de la informació que s'anomena la 'desigualtat del processament de dades' (*data processing inequality*) i deriva del fet que l'exemple assumeix que els estats formen una cadena de Markov. Per a més informació, vegeu Cover i Thomas (2006, 34-5).

inacceptable. De manera més general, el problema és que molts estats mentals porten més informació sobre causes pròximes que sobre els estats llunyans que els causen. Això s'anomena normalment 'el problema de la distància' i és una dificultat central de les teories informacionals.

Malgrat aquestes objeccions, les teories informacionals estan clarament en auge. Les més populars, però, segueixen sent les teories teleològiques o teleosemàntiques, que analitzarem a l'apartat següent.

2.4 Teories teleològiques

La darrera proposta que discutirem en detall és l'enfocament teleològic o teleosemàntic. A diferència de les teories que hem vist fins ara, una de les característiques principals de les propostes teleològiques és que normalment no només intenten oferir una resposta a la pregunta semàntica (per què un estat representa S en comptes de S*?), sinó també a la qüestió metasemàntica: què fa que un estat sigui una representació? Per què certs patrons d'activació neuronal o certes taques de tinta sobre un full de paper són representacions amb contingut i, en canvi, la posició dels núvols o un cop de vent no ho són? El fet d'abordar aquesta pregunta més fonamental incrementa significativament l'interès per les teories teleològiques, però també augmenta el nombre de reptes i dificultats que cal resoldre per poder considerar-la una teoria satisfactòria.

De manera simplificada, podríem dir que les teories teleològiques afirmen que la capacitat de representar és una habilitat que certs organismes han desenvolupat evolutivament, i per tant, la millor manera d'enfocar aquest problema és des del punt de la vista de la biologia (no és, per tant, sorprenent, que el llibre fundador d'aquest corrent tingui un títol tan provocador com '*Llenguatge, pensament i altres categories biològiques*'). Tot i que, com és obvi, hi ha diferents teories teleològiques i en aquest breu assaig no podem entrar a valorar-les en profunditat, en totes elles juga un paper central la noció de *funció*. Per aquest motiu, és important que l'examinem en detall.

2.4.1 Funció

L'eix vertebrador de les teories teleològiques és la noció de *funció*, tal com s'usa en biologia. És la noció que emprem quan diem que la funció del cor és bombar sang, la funció del pàncrees és produir insulina o la funció de la saliva és contribuir a la descomposició dels aliments. En general, les funcions són un tipus d'efectes. Els cors, per exemple, tenen molts efectes (bombar sang, fer sorolls bum-bum...), però només d'alguns d'ells diem que són les seves funcions.

Ara bé, què determina la funció d'un tret? Aquesta és una qüestió molt discutida en filosofia de la biologia i certament no hi ha una posició unànime (ARIEW *et al.* 2002; BULLER 1999). Tanmateix, una teoria àmpliament acceptada és l'anomenada teoria 'etiològica' de les funcions. Segons aquesta proposta, la funció d'un tret és l'efecte que explica perquè el tret va ser seleccionat evolutivament. Per què els organismes vius com nosaltres tenim cors? Tot i que aquest òrgan té molts efectes, no tots ells van contribuir a fer els nostres avantpassats més aptes. Per exemple, el fet que els cors fossin vermells o que generessin uns sorolls rítmics no va ajudar en res als organismes que tenien aquest òrgan. En canvi, el fet que els cors bombessin sang va ser crucial, per exemple, perquè l'oxigen arribés a les cèl·lules del cos. Per tant, aquest és l'efecte que va contribuir positivament a la supervivència dels nostres avantpassats i, per això, la funció del cor és bombar sang i no ser de color vermell o fer determinats sons. De manera general, la teoria etiològica diu que la funció d'un tret és l'efecte en virtut del qual el tret ha estat seleccionat per selecció natural.

La teoria etiològica té diversos avantatges des del punt del vista del filòsof naturalista. En primer lloc, d'acord amb aquesta proposta les funcions d'un tret no depenen en absolut de les intencions o els interessos dels científics i, per tant, té les credencials necessàries per a formar part d'una teoria naturalista de la intencionalitat (en canvi, altres definicions de funció la fan dependre dels estats intencionals de l'observador; CUMMINS 1975). En segon lloc, pot explicar per què les funcions són normatives (i, per tant, per què en última instància la intencionalitat també posseeix normativitat). Les funcions són propietats normatives. La funció d'un tret és quelcom que el tret *hauria de fer*. Potser no tots els cors bomben sang, però se suposa que tots ho haurien de fer perquè aquesta és la seva funció. Un cor que no bomba sang funciona *malament* o *erròniament* i és *disfuncional*. Això és un aspecte central a l'hora d'intentar explicar la intencionalitat perquè, com veiem, les representacions també són estats amb propietats normatives. Les representacions poden ser *falses* o *errònies*, i sembla que quan un mecanisme produeix representacions falses, hi ha alguna cosa que ha anat *malament*. Molta gent té l'esperança que, apel·lant a una noció normativa com és la de funció, podria ser possible explicar la normativitat dels estats intencionals i, a més, fer-ho a partir d'una noció més bàsica i acceptada per la ciències.

No obstant això, la noció de funció no pot ser l'únic element d'una teoria de la intencionalitat, per la senzilla raó que hi ha molts trets que tenen funcions però no són representacions. Ni el cor representa el bombeig de la sang, ni el pàncrees representa la insulina. El segon ingredient que cal per a ser una representació és un sistema emissor-receptor, que és un

model habitual en teoria de la comunicació. Un sistema emissor-receptor simplement consta d'un mecanisme que emet certs signes i un mecanisme que rep aquests signes i els interpreta. Pensem en un cas típic de comunicació humana: hi ha un agent que parla (l'emissor) i un agent que escolta (el receptor). Evidentment, un sistema emissor-receptor que pugui ser usat en una teoria naturalista no es pot definir a partir d'agents humans que tenen la intenció d'emetre sons o que empren signes, perquè aquestes descripcions pressuposen precisament el que estem intentant explicar, a saber, la intencionalitat. Tanmateix, podem definir i utilitzar un esquema emissor-receptor prescindint d'aquests elements: és un sistema compost per dos mecanismes, de tal manera que el primer rep certs estats externs com a input i produeix un estat intermedi com a output, i el segon mecanisme rep l'estat intermedi com a input i produeix un estat final com a output. D'alguna manera, podríem concebre el sistema emissor-receptor com un sistema completament automatitzat, en què un mecanisme produeix estats intermedis que l'altre rep i que condicionen la seva pròpia activitat.

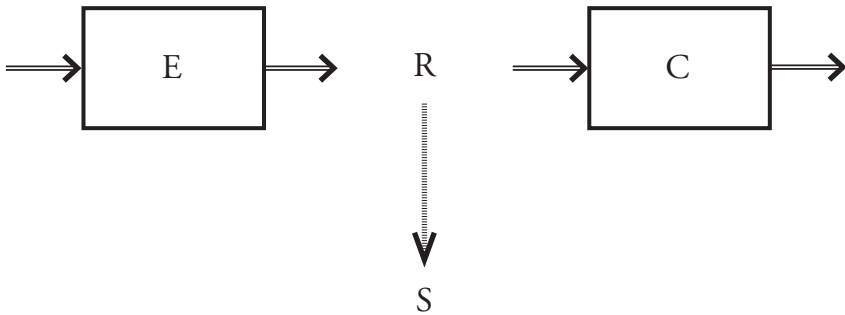


Figura 2: Sistema representacional, compost per un emissor (E) i un receptor (C). La funció de l'emissor E és produir una representació R quan S és el cas i la funció del receptor C és produir un efecte quan R és generat.

La idea fonamental de les teories teleològiques és que les representacions es poden definir a partir de les nocions de funció i sistema emissor-receptor. En particular, els sistemes representacionals són sistemes en els quals hi ha un emissor que té la funció de produir estats intermedis quan alguna cosa és el cas, i el receptor té la funció d'actuar de certa manera quan aquest estat intermedi és produït (figura 2). Les representacions no són res més que els estats intermedis produïts per emissors i rebuts per receptors, on el primer té la funció de produir-lo i el segon la funció d'actuar de certa manera quan l'estat intermedi és produït. Això és:

(TT) Un estat R és una representació de S si i només si:

- (1) R és produït per un mecanisme que té la funció de produir R quan S és el cas
- (2) R és rebut per un mecanisme que té la funció d'actuar quan R és produït

Per il·lustrar la proposta, considerem el famós cas del primat catari *Chlorocebus pygerythrus*. Com molts altres animals, aquests primats han desenvolupat un conjunt de signes d'alarma per avisar els seus companys d'un perill imminent. Quan un primat detecta un depredador alat que s'acosta (normalment una àliga) emet un grunyit de to baix, i els companys que el senten miren enlaire i corren a amagar-se als arbustos més pròxims. Si detecta una serp, emet un so breu i repetitiu i tots es posen drets per a poder observar millor el terra. Finalment, si albira un depredador mamífer terrestre (normalment un lleopard) emet uns crits secs i molt aguts i la resta de primats fugen i s'enfilen a l'arbre més pròxim, on un lleopard tindria moltes dificultats per a arribar. Segons la teoria teleològica, aquests sons emesos pel primat que descobreix el depredador són representacions: hi ha un emissor (el mico que veu el perill) que té la funció de produir un so (grunyit, so breu o crit) quan hi ha un perill (àliga, serp o lleopard) i hi ha un receptor (els altres micos) que tenen la funció d'actuar de certa manera quan reben el so. Per tant, cada un d'aquests sons són representacions i el seu contingut és *àliga, serp i lleopard a prop*, respectivament. Això sembla coincidir perfectament amb les nostres intuïcions i les dels biòlegs que descriuen aquests casos (CHENEY i SEYFARTH 1990).

Els sistemes emissor-receptor poden ser mecanismes en diferents animals (com el cas anterior, en què un primat emet un so i un altre l'escolta), però també poden ser diferents parts d'un mateix organisme. Les libèl·lules, per exemple, tenen un sistema de caça molt efectiu. Quan detecten una ombra de certa mida (al voltant dels 2º) desplaçant-se a certa velocitat, es disparen un conjunt de neurones anomenades '*Target-Selective Descending Neurons*' (TSDN) que activen el sistema motor i dirigeixen la libèl·lula a interceptar la presa. Sorprenentment, en un 95% dels casos aconsegueixen capturar-la (COMBES *et al.* 2013; GONZÁLEZ-BELLIDO *et al.* 2013, 699; OLBERG *et al.* 2000, 155). També en aquest cas sembla que la teoria teleològica prediu encertadament que l'activació de les neurones TSDN és un representació: hi ha un mecanisme que té la funció d'activar les neurones TSDN quan hi ha una presa desplaçant-se en certa direcció, i hi ha un mecanisme que té la funció de produir un comportament molt determinat quan aquest estat intermedi es produït. En aquest cas, l'emissor seria el sistema perceptiu, el receptor el sistema motor i la representació l'estat neuronal. Per tant, d'acord amb les teories

teleològiques, aquest estat cerebral és una representació i el seu contingut és, aproximadament, *presa a prop*.

En conseqüència, les teories teleològiques ofereixen una solució a la qüestió metasemàntica: un estat és una representació si i només si és un estat produït per un sistema emissor-receptor que té certes funcions. A més a més, també dona un criteri per a determinar el contingut de la representació que sembla solucionar els problemes de les teories anteriors. D'una banda, deixen lloc a l'error: de la mateixa manera que les funcions són efectes que els trets haurien de tenir, però que de vegades no tenen, el mecanisme que té la funció de produir grunyits quan hi ha depredador alat, en una ocasió concreta pot produir el signe quan de fet no n'hi ha cap (per exemple, quan passa una gavina inofensiva). En aquest cas, el mecanisme no estarà complint la funció, i direm que la representació és falsa. En altres paraules, podem explicar com poden haver-hi estats intencionals erronis apel·lant a l'error associat a les funcions biològiques. De fet, TT fins i tot és compatible amb l'error freqüent: així com la funció dels espermatozoides és fertilitzar un òvul i només un entre milions ho aconsegueix, pot ser que la funció del grunyit sigui indicar la presència d'una àliga, però que la majoria de vegades aquesta representació sigui falsa. Sabem del cert que hi pot haver selecció d'animals porucs (GODFREY-SMITH 1991).

Malauradament, no tot són virtuts. Si bé durant els últims anys les teories teleològiques s'han desenvolupat i perfeccionat en molts aspectes, també s'han presentat moltes objeccions. Tot seguit en discutirem algunes de les més importants.

Indeterminació

Un primer problema clàssic que ja hem analitzat és la indeterminació del contingut. Pensem de nou en el cas de les libèl·lules. Dèiem que la teoria teleològica sembla implicar que el seu estat cerebral representa *presa a prop*, perquè és plausible pensar que el mecanisme neuronal que és responsable de l'activació de les neurones TSDN té la funció de produir-la quan hi ha una presa a prop. Tanmateix, algú podria suggerir que la teoria teleosemàntica prediu un contingut diferent: *cosa negra movent-se*. És a dir, la funció del sistema perceptiu podria ser produir una activació neuronal quan hi ha una cosa negra que es desplaça, i no una presa. Com que en el passat un nombre suficientment alt de coses negres que es desplaçaven eren preses, representar una cosa negra i llençar-s'hi a sobre pot ser un procediment tan efectiu com representar una presa i llençar-s'hi (per a una demostració matemàtica, vegeu MARTÍNEZ 2013A). De fet, els problemes no s'acaben aquí: el contingut també podria ser *insecte volador, cosa amb certes proteïnes, punt negre a la retina...* De nou, aquesta teoria

pateix el problema de la indeterminació perquè atribueix continguts múltiples en casos en què la ciència i el sentit comú atribuirien un únic contingut.¹⁸

Hom podria estar temptat a menystenir aquest problema en el cas dels micos o les libèl·lules i simplement acceptar que aquests organismes cognitivament tan poc sofisticats tenen representacions amb continguts indeterminats (Papineau, 1998). Amb tot, aquest resultat té costos importants. En primer lloc, cal destacar que amb aquesta indeterminació és molt difícil assegurar el valor explicatiu de les representacions. Tal com hem vist al principi de l'article, l'atribució de representacions té un paper central en moltes explicacions científiques i no científiques. En concret, si volem explicar per què la libèl·lula surt disparada en certa direcció quan s'activen les neurones TSDN, sembla que hem d'apel·lar al fet que aquesta activació representa *presa a prop*. En canvi, si l'estat cerebral representés *punt negre a la retina*, seria molt difícil entendre el seu comportament (per què fugir quan hi ha un punt negre a la retina?). Per tant, un contingut tan indeterminat està en tensió amb el rol explicatiu que tenen les atribucions de contingut. En segon lloc, recordem que tot i que ens estem centrant en animals no humans, l'objectiu d'aquesta teoria és ser capaç de naturalitzar tots els estats intencionals, dels més simples als més complexes. Ara bé, si acceptem aquest grau d'indeterminació tant gran respecte als d'insectes, és difícil veure de quina manera aquesta teoria podrà predir continguts més determinats per molts dels nostres estats cerebrals subpersonals¹⁹ (que segurament comparteixen moltes propietats amb les neurones de les libèl·lules) o fins i tot representacions més complexes com els conceptes, el pensament o les imaginacions. Per exemple, si admeten la indeterminació del contingut en aquest estadi, no és obvi de quina manera podrem justificar que el nostre concepte GOS refereix a *gos* i no *animal pelut amb quatre potes*, o *taca de color de mida mitjana*, etc.

Evidentment, s'han presentat diferents respostes a aquesta objecció. Alguns accepten la indeterminació i s'esforcen a mostrar que no és tan problemàtica com sembla (PAPINEAU 1998; AGAR 1993). D'altres han intentat argumentar que una noció adequada de funció biològica pot solucionar-ho

¹⁸ De vegades s'empra el nom 'problema de la disjunció' (*Disjunction Problem*) per referir-se al problema de la indeterminació, i d'altres s'usa per referir-se al problema de l'error que hem destacat abans. El fet d'utilitzar la mateixa expressió per dos fenòmens diferents sovint porta a confusions (FODOR, 1990), i per això aquí he obviat aquest nom.

¹⁹ És habitual distingir entre estats personals i estats subpersonals, tot i que el criteri per a distingir-los és controvertit. Estats personals paradigmàtics són les creences o els desitjos i, com el seu nom indica, són atribuïbles a la persona (per exemple, el Joan creu o desitja que p). En canvi, un exemple d'estat subpersonal és l'activació de l'Àrea Facial Fusiforme que fem en el reconeixement de cares (vegeu la secció 1). Normalment, els estats subpersonals no són accessibles a la consciència i no s'atribueixen a la persona com a tal (BERMÚDEZ 2005, 27-31).

(MILLIKAN 2005; NEANDER 2013; PRICE 1998). En canvi, alguns autors han cregut oportú complementar la teoria teleològica amb alguna condició addicional que permeti reduir la indeterminació (MARTÍNEZ, 2013A). Sigui com sigui, aquest problema segueix essent una de les dificultats més importants de les teories teleològiques.

L'home del pantà

Com hem vist abans, a diferència d'altres propostes, les teories teleològiques no només ofereixen una resposta a la qüestió semàntica (això és, explicar què fa que les representacions tinguin un contingut determinat) sinó també a la metasemàntica (explicar què es una representació). Aquest és un projecte molt més interessant i ambiciós, però evidentment també origina noves objeccions. L'home del pantà és segurament la més coneguda.

Imagineu que un dia en Joan Monleon passeja pel bosc quan de sobte un llamp li cau directament al cap. El que s'esdevé és sorprenent: si bé en Joan Monleon es desintegra completament, el mateix llamp produeix (per una coincidència gairebé màgica) un duplicat exacte d'en Joan Monleon a partir de reorganitzar les molècules contingudes en una bassa que hi ha al costat del camí. Evidentment, la creació d'aquest 'home del pantà' (*Swampman*, en anglès) és altament improbable, però no sembla que aquest esdeveniment sigui impossible metafísicament o conceptual. El que és desconcertant d'aquest cas és que segons la teoria etiològica en Joan del Pantà no té representacions mentals; en Joan del Pantà no té una història evolutiva (no és un humà!), els seus òrgans no han estat seleccionats per l'evolució i per tant, segons la teoria etiològica, cap dels seus trets no té funció. Pròpiament parlant, dins del seu cap no té un cervell, sinó una còpia que per atzar s'assembla extremadament i en tot detall a un cervell de veritat. Ara bé, si els trets d'en Joan del Pantà no tenen funcions, llavors segons la teoria teleològica tampoc podran produir representacions. Per tant, una conseqüència de les teories teleològiques és que un ésser sense història (per exemple, un ésser creat per atzar per un llamp) no pot tenir representacions.

És important remarcar com n'és de contraintuïtiva aquesta conclusió. En Joan del Pantà és exactament igual, molècula per molècula, a en Joan Monleon, però un té representacions i l'altre no, perquè li manca l'adequada història evolutiva. Suposem que, després de la seva creació, en Joan del Pantà s'aixeca una mica estabornit, es dirigeix cap a casa en Joan Monleon i segueix la vida que hauria fet ell si hagués estat viu: fa teatre, grava discos, presenta programes de televisió, etc. Intuïtivament, si necessitem atribuir representacions per a explicar el comportament d'en Joan Monleon, també les necessitaríem per explicar el comportament d'en Joan del Pantà. Però segons la teoria etiològica,

en Joan del Pantà no en té cap ni una perquè li manca la història adequada. L'home del pantà és un greu problema per la teoria etiològica.

Hi ha diverses maneres d'enfrontar-se a aquest problema. En primer lloc, es podria defensar que tot i que en Joan del Pantà hagi estat produït fortuïtament, la seva semblança amb nosaltres fa que pugui ser classificat com a humà, i per tant els seus òrgans també tinguin la categoria d'òrgans humans. D'aquesta manera, dins del cap tindria un cervell amb les mateixes funcions biològiques que el nostre. Tanmateix, perquè aquesta proposta no sigui una solució *ad hoc*, necessitariem explicar i justificar un criteri general per a classificar trets biològics que tingués aquest resultat (una tasca gens trivial). Una solució alternativa seria acceptar que en Joan del Pantà no té representacions al principi, però que quan comença a interactuar amb els altres i amb el món, els seus trets van adquirint funcions i per tant, progressivament va adquirint representacions. Això requeriria explicar amb molt detall quines són aquestes funcions i com es poden establir (no està clar, per exemple, que això sigui compatible amb la teoria etiològica). En tercer lloc, hom podria acceptar aquesta conseqüència (que en Joan del Pantà no té representacions ni les tindrà mai perquè els seus trets no són producte de la selecció natural), però pot intentar fer aquesta conclusió menys inversemblant (DRETSKE 1995; MILLIKAN 1993). Per exemple, molts filòsofs creuen que per a poder tenir pensaments sobre persones individuals (la pròpia mare, fill, etc.) o substàncies naturals (aigua, or,..) és necessari que el subjecte hagi tingut una relació causal amb aquestes coses (KRIPKE 1980; PUTNAM 1975). Si això és així, llavors quan en Joan del Pantà és creat no pot tenir pensaments sobre l'aigua, ni sobre els avets. Això sembla contraintuïtiu, però és una conseqüència directa d'una de les teories més populars de la referència conceptual. Per tant, si molts filòsofs ja estan compromesos amb el fet que en Joan del Pantà no té certs estats mentals, només falta un petit pas perquè també acceptin que, de fet, tampoc té cap altra representació.

Pansemanticisme

En aquest punt m'agradaria destacar una altra objecció que ha ressorgit amb força recentment i que, d'alguna manera, constitueix el pol oposat al problema anterior. Si bé l'objecció de l'home del pantà planteja que les teories teleològiques segurament prediuen que no hi ha representacions en casos en què intuïtivament sí que n'hi ha, alguns filòsofs han argumentat que en d'altres casos les teories teleològiques prediuen l'existència de representacions on intuïtivament no n'hi ha. Per exemple, molts bacteris tenen mecanismes per a produir molècules que són emeses a l'exterior de la membrana cel·lular i que són captades per altres bacteris a través de certs receptors. Com que hi ha un emissor i un receptor que han evolucionat per selecció natural (i que,

per tant, tenen funcions), segons les teories teleològiques aquestes molècules emeses haurien de ser qualificades de representacions. Un cas similar és de les hormones. Quan augmenta l'osmolaritat de la sang, la glàndula pituïtària (situada a la part inferior del cervell) allibera una hormona anomenada 'arginina vasopressina'. Aquesta hormona viatja a través de la sang i activa uns receptors a diferents parts del cos que, entre altres coses, fan augmentar la reabsorció d'aigua (la qual cosa redueix l'osmolaritat). Schulte (en premsa) argumenta que la teoria teleològica implica que l'arginina vasopressina és una representació, ja que hi ha emissors i receptors amb funcions. Sterelny (1995) mostra que segons les teories teleològiques la saliva també seria una representació de menjar, ja que és produïda per la glàndula salival quan detectem menjar i és rebuda per l'estómac, i ambdós mecanismes tenen funcions biològiques. En resum, sembla que la teoria teleològica és excessivament liberal a l'hora d'atribuir representacions. Recentment, aquesta objecció ha estat desenvolupada per un nombre significatiu de filòsofs (BURGE 2012; RESCORLA 2013; RAMSEY 2007).

De nou, davant d'aquesta dificultat el teleosemanticista té diverses opcions (ARTIGA, en premsa). En primer lloc, pot acceptar que tots aquests exemples són realment representacions i pot intentar fer aquesta conclusió més plausible. Millikan (2004) segueix aquesta estratègia i argumenta que si bé la teleosemàntica implica que algunes molècules emeses pels bacteris són representacions, són casos marginals; de la mateixa manera que el 0 és un número, tot i que dista molt de ser-ne un de paradigmàtic, els bacteris emeten representacions tot i no ser un cas central. De fet, en alguns casos, acceptar que siguin representacions sembla bastant plausible. En concret, molts microbiòlegs parlen de 'comunicació bacteriana' i assumeixen que són signes de ple dret (MILLER i BASSLER 2001; WITZANY 2011). També les hormones són moltes vegades classificades com a signes que viatgen a través de la sang (ARANDA 2015). Per tant, caldria veure quins d'aquests exemples presenten realment una objecció. La segona estratègia (que és plenament compatible amb la primera) és afegir algun criteri més que un estat ha de satisfer per a poder tenir l'estatus de representació. Per exemple, alguns han plantejat que les representacions són estats que tenen una relació arbitrària amb el seu contingut (STEGMANN 2004).

Representacions complexes

Finalment, és important assenyalar que la complexitat de certes representacions presenta un repte afegit. L'esquema bàsic de les teories teleològiques assumeix que hi ha un emissor i un receptor, que l'emissor té la funció de produir un estat concret quan hi ha un estímul específic (presa, vermell...) i que el receptor té la funció de generar un comportament determinat. Això certament encaixa amb alguns dels exemples que hem discutit: la resposta de les libèl·lules

a una presa és força automàtica i sempre que una ombra de certa mida es desplaça a una determinada velocitat per una àrea del seu camp visual, la libèl·lula sortirà disparada a capturar-lo. És obvi, però, que moltes altres representacions no estan lligades tan estretament a cap estímul concret. Per exemple, molts de nosaltres som capaços de reconèixer les nostres mascotes veient-les des de diferents angles, quan són parcialment amagades darrere un sofà, a molta distància, o fins i tot per l'olor o el so que emeten. Igualment, moltes de les nostres representacions no produeixen sempre la mateixa acció; quan veiem la nostra mascota, podem realitzar diverses accions: de vegades la cridem, l'abracem, la renyem o, simplement, la ignorem. Incorporar sistemes representacionals més complexos dins del patró assumit per la teleosemàntica és tot un repte.

De fet, algunes de les propietats que tenen les representacions complexes no semblen encaixar en el marc de les teories teleològiques. Per exemple, el nostre llenguatge (i probablement el nostre pensament —FODOR 1975—) és composicional, això és, hi ha un conjunt de regles que estableixen de quina manera podem formar representacions més complexes a partir dels seus elements més simples. També podem crear noves paraules i conceptes que no havien existit abans —‘neutri’, ‘ordinador’, ‘supernova’, etc.—. No està clar de quina manera una teoria teleològica que apel·la a la història evolutiva pot incorporar tots aquests fenòmens. Evidentment, això no és tant una objecció com un àmbit on segurament les teories teleosemàntiques poden trobar dificultats. Recentment s'ha posat molt èmfasi en aquesta qüestió i, entre d'altres coses, s'ha discutit de quina manera les teories teleològiques poden aplicar-se a nivell neuronal (RYDER 2006; CAO 2012; ARTIGA 2016) i s'ha abordat la qüestió de la productivitat (MARTÍNEZ 2013B). Esperem veure nous desenvolupaments en aquesta direcció ben aviat.

Al problema de la indeterminació, l'home del pantà, el pansemanticisme i les representacions complexes se'n podrien afegir alguns més (PIETROSKI 1992; MENDELOVICI 2013). Sigui com sigui, és un debat plenament actual, en què hi ha moltes qüestions obertes i molt de camp per recórrer (MACDONALD i PAPINEAU 2006; NEANDER 2012; SHEA 2013B).

2.5 Altres opcions

Fins ara hem considerat diverses teories naturalistes del contingut (teories de la semblança, causals, informacionals i teleològiques) i hem vist que, malgrat ser les propostes més prometedores, totes s'enfronten a objeccions importants. Arribats a aquest punt, algú podria pensar que les dificultats són tan serioses que el més probable és que cap d'elles no aconsegueixi solucionar-les totes. En aquesta darrera secció voldria discutir breument les opcions

disponibles per aquelles persones que estiguin temptades a rebutjar totes les teories anteriors: l'eliminativisme, el ficcionalisme, i el pluralisme.²⁰

En primer lloc, algú podria argumentar que el fracàs de totes aquestes teories mostra que realment no hi ha representacions. Aquesta postura tan extrema, anomenada *eliminativisme* respecte a les representacions, ha estat defensada en certes àrees, sobretot a l'àmbit de les ciències cognitives i la intel·ligència artificial (CHURCHLAND 1981; BROOKS 1991; O'REAGAN i NOE 2001; HUTTO i MYIN 2013). Tanmateix, com vèiem al principi, sembla que la noció de representació juga un paper indispensable en moltes de les explicacions científiques i no científiques. Per exemple, intentar donar una explicació psicològica del comportament humà sense apel·lar a representacions perceptives, experiències, desitjos, creences, conceptes o alguna altra representació sembla del tot impossible.

Una postura alternativa defensada recentment per alguns teòrics i relativament menys extrema és el *ficcionalisme* respecte a les representacions, que sosté que tot i que realment no hi ha representacions, postular-ne l'existència ens és molt útil per certs projectes explicatius (LEVY 2011; SPREVAK 2013). Pensem en el cas de la literatura: quan diem que *la Colometa* vivia a Barcelona estem dient una cosa certa dins de la ficció (en particular, de *La Plaça del Diamant* de Mercè Rodoreda), però fora d'aquest context sabem que realment la Colometa no va existir i que, per tant, no vivia a Barcelona (ni enlloc). De la mateixa manera, segons aquesta postura quan els científics elaboren una explicació en la qual apel·len a estats representacionals, potser allò que diuen és cert dins de la ficció, però no es comprometen a què existeixin fora d'ella. El ficcionalisme és una postura molt atractiva a primera vista, perquè sembla combinar el millor dels dos mons: d'una banda, manté el poder explicatiu de l'atribució de representacions, però de l'altra l'allibera de la responsabilitat d'haver de donar-ne una teoria realista. Tanmateix, aquesta postura també té greus problemes. En primer lloc, sembla que hi ha una tensió entre el gran poder explicatiu que té pressuposar representacions i la seva inexistència. Si realment necessitem suposar que hi ha representacions per explicar certs fenòmens, i aquest postulat és tan útil, per què això no és un argument poderós a favor de la seva existència? De fet, en general una de les raons més convincentes a favor l'existència de qualsevol entitat (els àtoms, la selecció natural, etc.) és la gran capacitat explicativa que té suposar-la. Semblaria gairebé un miracle que una entitat no existeixi i que, tanmateix, pressuposar-la

²⁰ Una altra opció és defensar que la noció de 'representació' és bàsica i no es pot analitzar de cap manera (una posició que podríem anomenar 'no naturalista'). Tanmateix, això simplement sembla abandonar el projecte, més que oferir-ne una solució. Aquesta proposta no resol ni clarifica cap de les propietats aparentment misterioses que hem apuntat al principi i, per tant, no ens ajuda a entendre aquesta noció tan bàsica per a la ciència i el sentit comú.

sigui extremadament útil i fins i tot imprescindible en certes explicacions.²¹ La segona dificultat del ficcionalisme és encara més greu; si realment no existeixen les representacions, llavors la mateixa teoria no es pot entendre, ni pensar, ni acceptar, perquè la comprensió, el pensament o l'acceptació són representacions mentals (SPREVAK 2013). Per tant, el mer fet que es pugui acceptar o rebutjar la teoria, ja sembla refutar-la.

Hem vist que l'eliminativisme i el ficcionalisme respecte a les representacions són opcions difícils de sostenir. La darrera opció per tot aquell que cregui que cap de les teories anteriors pot solucionar tots els problemes és adoptar alguna forma de pluralisme. Potser donar una explicació general que sigui adequada per a totes les representacions no és possible i, per tant, es podria sostenir que cada una de les teories que hem presentat abans és veritat només respecte a un tipus d'estat intencional (justament en aquell en què no tingui problemes greus). Per exemple, la teoria de la semblança seria veritat respecte a les representacions icòniques (imatges, pintures, etc.) i la teoria teleològica respecte als estats mentals dels insectes. No obstant això, qui volgués defensar aquesta opinió hauria d'enfrontar-se a dues qüestions importants. En primer lloc, defensar un pluralisme no eludeix la responsabilitat de donar una teoria adequada dels diferents tipus de representacions. És a dir, si creiem que la teoria de la semblança és adequada per a les representacions icòniques i la teoria informacional per als estats perceptius, caldria elaborar amb detall una teoria adequada per a aquestes teories en aquests àmbits. Atès el gran nombre de dificultats que cada una d'aquests propostes té, això tendria a multiplicar els problemes, més que oferir-ne una solució. En segon lloc, si creiem que els estats intencionals (creences, imatges, estats cerebrals dels insectes, etc.) són representacions en virtut de processos diferents, caldria preguntar-se si tenen res de significatiu en comú que justifiqui classificar-los conjuntament dins del mateixa categoria. Si no hi ha cap propietat interessant que totes les representacions comparteixin, es podria pensar que, més que una teoria naturalista, el que hom defensa és un eliminativisme respecte de les representacions. En altres paraules, s'estaria suggerint que usem aquesta expressió per referir-nos a un conjunt de fenòmens diversos que no tenen en comú cap tret suficientment significatiu per a agrupar-los dins de la mateixa classe. Per tant, el pluralisme podria heretar les mateixes dificultats que l'eliminativisme.

Per tant, tot i que les teories naturalistes de la intencionalitat han de front a objeccions serioses, cal emfatitzar que qualsevol proposta alternativa (eliminativisme, ficcionalisme, pluralisme) també té dificultats pròpies. Això

²¹ Efectivament, això és una versió del que en filosofia de la ciència s'anomena '*No-Miracles Argument*' i que s'ha emprat a bastament per a defensar el realisme científic.

mostra que en el debat sobre una noció tan central com la intencionalitat no hi ha cap posició còmoda que es pugui sostenir sense haver de fer front a un conjunt important d'arguments.

3. CONCLUSIONS

Tot i que en les últimes dècades hem avançat significativament en la comprensió i l'anàlisi dels estats intencionals, sembla que encara som lluny de poder fer front als grans reptes que plantegen aquestes entitats. D'una banda, necessitem apel·lar a les representacions tant en el dia a dia com en les nostres millors explicacions científiques i filosòfiques, però, de l'altra, semblen ser entitats desconcertants. Encara no sabem de quina manera els estats intencionals són compatibles amb una visió naturalista de l'univers. I, tanmateix, el debat sobre la naturalització de la intencionalitat mostra clarament que el treball conjunt de la filosofia i la ciència és imprescindible per avançar significativament en una millor comprensió de nosaltres mateixos i del món que ens envolta.

Agraïments: Vull agrair al David Pineda, al Sergi Rosell i a la Verònica Zaragoza Gómez els seus suggeriments i comentaris. Aquest treball ha estat possible gràcies a la beca de formació postdoctoral FPDI-2013-16764 del Ministeri d'Economia i Competitivitat i el projecte "La Complejidad de la Percepción: Un Enfoque Multidimensional" (FFI2014-51811-P).

BIBLIOGRAFIA

- ADAMS, F. i AIZAWA, K. 1992, “‘X’ Means X: Semantics Fodor-Style”, *Minds and Machines*, 2: 175-83.
- ADAMS, F. i AIZAWA, K. 2010, “Causal Theories of Mental Content”, *Stanford Encyclopedia of Philosophy*.
- AGAR, N. 1993, “What Do Frogs Really Believe?”, *Australasian Journal of Philosophy*, 71 (1): 1§12.
- ARANDA, A. 2015, *Las hormonas*, CSIC Ediciones.
- ARTIGA, M. 2014, “Prinz’s Naturalistic Theory of Intentional Content”, *Crítica*, 46 (136): 69-86.
- ARTIGA, M. 2016, “Teleosemantic Modeling of Cognitive Representations”, *Biology and Philosophy*, 31 (4): 483-505.
- ARTIGA, M. (en premsa), “Liberal Representationalism: A Deflationist Defense”, *Dialectica*.
- ARIEW, A.; CUMMINS, R. i PERLMAN, M. (eds.) 2002, *Functions: New Essays in the Philosophy of Psychology and Biology*, Oxford University Press.
- BAIN, D. T. 2003, “Intentionalism and pain”, *Philosophical Quarterly*, 53 (213): 502-23.
- BECHTEL, W. 2016, “Investigating Neural Representations: The Tale of Place Cells”, *Synthese* 193 (5): 1287-321.
- BERMÚDEZ, J. L. 2005, *Philosophy of Psychology. A Contemporary Introduction*, Routledge.
- BIRCH, J. 2014, “Propositional Content in Signalling Systems”, *Philosophical Studies*, 171 (3): 493-512.
- BOCCARA, C. N.; SARGOLINI, F.; THORESEN, V. H.; SOLSTAD, T.; WITTER, M. P.; MOSER, E. I. i MOSER, M.-B. 2010, “Grid cells in pre- and parasubiculum”, *Nature Neuroscience*, 13: 987-94.
- BROOKS, R. 1991, “Intelligence Without Representation”, *Artificial Intelligence*, 47: 139-59.
- BULLER, D. (ed.) 1999, *Function, Selection, and Design*, SUNY Press.
- BURGE, T. 2010, *The Origins of Objectivity*, Oxford University Press.
- CAO, R. 2012, “A teleosemantic approach to information in the brain”, *Biology and Philosophy*, 27 (1): 49-§71.
- COMBES, S.; SALCEDO, M.; PANDIT, M. i IWASAKI, J. 2013, “Capture success and e-ciency of dragonies pursuing dierent types of prey”, *Integrative and Comparative Biology*, 53 (5):787§-798.
- COVER, T. M. i THOMAS, J. A. 2006, *Elements of InInformation Theory*, Wiley-Interscience.

- CHENEY, D. L. i SEYFARTH, R. M. 1990, *How monkeys see the world: inside the mind of another species*, University of Chicago Press.
- CHURCHLAND, P. 1981, Eliminativist Materialism and Propositional Attitudes, *Journal of Philosophy*, 78 (2): 67-90.
- CHURCHLAND, P. 2001, "Neurosemantics: On the mapping of minds and the portrayal of worlds", en K. E. WHITE (ed.), *The emergence of mind*, 117-47.
- COVER, T. M.; THOMAS, J. A. 2006, *Elements of Information Theory*, Wiley-Interscience.
- CUMMINS, R. 1975, Functional Analysis, *Journal of Philosophy*, 72: 741-64.
- CUMMINS, R. 1996, *Representations, targets, and attitudes*, The MIT Press.
- DRETSKE, F. 1995, *Naturalizing the Mind*, MIT Press.
- ELIASMITH, C. 2000, *How neurons mean: A neurocomputational theory of representational content*. Tesi Doctoral, Washington University in St. Louis.
- ELIASMITH, C. 2005, "A new perspective on representational problems". *Journal of Cognitive Science*, 6: 97§-123.
- FLORIDI, L. 2010, *Information: A Very Short Introduction*, Oxford University Press.
- FODOR, J. 1990, *A Theory of Content and Other Essays*, The MIT Press.
- GONZALEZ-BELLIDO, P.; PENG, H.; YANG, J.; GEORGOPOULOS, A. i OLBERG, R. 2013, "Eight pairs of descending visual neurons in the dragonfly give wing motor centers accurate population vector of prey direction", *Proceedings of the National Academy of Sciences*, 110 (2): 696§-701.
- HAXBY, J.; HOMAN, E. i GOBBINI, M. 2002, "The distributed human neural system for face perception", *Trends in Cognitive Sciences*, 4 (6): 223-33.
- HUBEL, D. H. i WIESEL, T. N. 1959, "Receptive fields of single neurones in the cat striate cortex", *Journal of Physiology*, 148: 574-§591.
- HUTTO, D. i MYN, E. 2013, *Radicalizing Enactivism*. MIT Press.
- ISHAI, A. 2008, "Let's face it: it's a cortical network", *NeuroImage*, 40 (2): 415-9.
- FAIRGALL, S.; ISHAI, A. 2007, "Effective connectivity within the distributed cortical network for face perception", *Cerebral Cortex*, 17 (10): 2400-6.
- FLORIDI, L. 2010, *Information: A Very Short Introduction*, Oxford University Press.
- FODOR, J. 1975, *The Language of Thought*, Harvard University Press.
- FODOR, J. 1990, *A Theory of Content and Other Essays*, MIT Press.
- GODFREY-SMITH, P. 1991, "Signal, Detection, Action", *Journal of Philosophy*, 88 (12): 709-22.
- GRICE, H. P. 1961, "The Causal Theory of Perception", *Proceedings of the Aristotelian Society, Supp.*: 121-5.
- JACOB, P. 2014, "Intentionality", *Stanford Encyclopedia of Philosophy*.
- KANWISHER, N. i BARTON, J. 2010, "The functional architecture of the face system: integrating evidence from fmri and patient studies", dins J. HAXBY,

- A. CALDER, G. RHODES i M. JOHNSON (eds.), *Handbook of Face Perception*, Oxford University Press.
- LEVY, A. 2011, "Information in Biology: a Ficitonalist Account", *Noûs*, 45 (4): 640-57.
- LOCKE, J. 1690/1990, *An Essay Concerning Human Understanding*, Clarendon Press.
- LOWER, B; REY, G. 1991, *Meaning in Mind: Fodor and His Critics*, Blackwell Publishers.
- R. OLBERG 2012, "Visual control of prey-capture ight in dragonies", *Current Opinion in Neurobiology*, 22: 267§-271.
- JACOB, P. 2014, "Intentionality", *Stanford Encyclopedia of Philosophy*.
- KRIPKE, S. 1980, *Naming and Necessity*, Harvard University Press.
- NEANDER, K. 2012, "Teleological Theories of Mental Content", *Stanford Encyclopedia of Philosophy*.
- MACDONALD, G. i PAPINEAU, D. 2006, *Teleosemantics*, Oxford University Press.
- MARTÍNEZ, M. 2013A, "Teleosemantics and Indeterminacy", *Dialectica*, 67 (4): 427-53.
- MARTÍNEZ, M. 2013B, "Teleosemantics and Productivity", *Philosophical Psychology*, 26 (1): 47-68.
- MENDELOVICI, A. 2013, "Reliable Misrepresentation and Tracking Theories of Mental Representation", *Philosophical Studies*, 165 (2): 421-43.
- MILLER M. i BASSLER, B. 2001, "Quorum sensing in bacteria", *Annual Review of Microbiology*, 55: 165-99.
- MILLIKAN, R. 1984, *Language, Thought and Other Biological Categories*, MIT Press.
- MILLIKAN, R. 1989, "In Defense of Proper Functions", *Philosophy of Science*, 56 (June): 288-302.
- MILLIKAN, R. 1993, *White Queen Psychology and Other Essays for Alice*, The MIT Press.
- MILLIKAN, R. 2004, *Varieties of Meaning*, MIT Press.
- MILLIKAN, R. 2005, *Language: A Biological Model*, Oxford Univesity Press.
- MOSER, E.; KROPPF, E. i MOSER, M. 2008, "Place Cells, Grid Celle and the Spatial Representation System", *Annu Rev. Neurosci.*, 31: 69-89.
- NEANDER, K. 2012, "Teleological theories of mental content", *Standford Encyclopedia of Philosophy*.
- NEANDER, K. 2013, "Toward an informational teleosemantics", en D. RYDER, J. KINGSBURY i K. WILLIFORD (eds.), *Millikan and her Critics*, Wiley-Blackwell.
- O'REGAN, J. i A. NOË, A. 2001, "A Sensorimotor Account of Vision and Visual Consciousness", *Behavioral and Brain Sciences*, 24 (5): 883-917.
- PAPINEAU, D. 1998, *Australasian Journal of Philosophy*, 76 (1): 1-14.

- PURVES, D.; AUGUSTINE, G.; FITZPATRICK, D.; HALL, W.; LAMANTIA, A. i WHITE, L. 2008, *Neuroscience*, Sinauer Associates.
- PIETROSKI, P. 1992, "Intentionality and Teleological Error", *Pacific Philosophical Quarterly*, 73 (3): 267-82.
- PINEDA, D. 2012, *La Mente Humana. Introducción a la Filosofía de la Psicología*, Cátedra.
- PRICE, C. 1998, "Determinate Functions", *Noûs* 32 (1): 54-75.
- PRINZ, J. 2000, "The Duality of Content", *Philosophical Studies*, 100 (1): 1-34.
- PRINZ, J. 2004, *Gut Reactions: A Perceptual Theory of Emotion*, Oxford University Press.
- PUTNAM, H. 1975, "The Meaning of 'Meaning'", *Minnesota Studies in the Philosophy of Science*, 7: 131-93.
- RAMSEY, W. 2007, *Representation Reconsidered*, Cambridge University Press.
- RESCORLA, R. A. 2013, "Millikan on honeybee navigation and communication", en D. RYDER, J. KINGSBURY i K. WILLIFORD (eds.), *Millikan and her Critics*, Blackwell.
- RODRÍGUEZ-PEREYRA, G. 2014, "El Nominalisme en Metafísica", *Quaderns de Filosofia*, 1 (1): 13-35.
- RUPERT, R. 1999, "The Best Test Theory of Extension: First Principle(s)", *Mind and Language* 14 (3): 321-55.
- RYDER, D. 2006, "On Thinking of Kinds: A Neuroscientific Perspective", MACDONALD i PAPINEAU (eds.), *Teleosemantics*, Oxford University Press.
- SCARANTINO, A. 2015, "Information as a Probabilistic Difference Maker", *Australasian Journal of Philosophy*, 93 (3): 419-43.
- SCHIELDS, CH. 2016, "Aristotle's Psychology", *Stanford Encyclopedia of Philosophy*.
- SCHULTE, P. (en premsa), "Perceptual Representations: A Teleosemantic Answer to the Breadth-of-Application Problem", *Biology and Philosophy*.
- SHEA, N. 2013A, "Millikan's Isomorphism Requirement", dins *Millikan and Her Critics*, D. RYDER, J. KINGSBURY i K. WILLIFORD (eds.), Wiley-Blackwell, 63-80.
- SHEA, N. 2013B, "Naturalising representational content". *Philosophy Compass*, 8 (5): 496-509.
- SKYRMS, B. 2010, *Signals: Evolution, Learning, and Information*, Oxford University Press.
- SPREVAK, M. 2013, "Fictionalism About Neural Representations", *The Monist*, 96 (4): 539-60.
- STEGMANN, U. 2004, "The Arbitrariness of the Genetic Code", *Biology and Philosophy*, 19: 205-22.
- STERELNY, K. 1995, "Basic Minds", *Philosophical Perspectives*, 9: 251-70.
- STERNBERG, E. i STERNBERG, K. 2009, *Cognitive Psychology*, Wadsworth.

TYE, M. 2000, *Consciousness, Collor and Content*, MIT Press.

USHER, M. 2001, "A statistical referential theory of content: Using information theory to account for misrepresentation", *Mind and Language*, 16 (3): 331-334.

WITZANY, G. (ed.) 2011, *Biocommunication in Soil Microorganisms*, Springer.