Revista Semestral ISSN: 0798-4324 Depósito Legal: pp198102DF423

# EPISTEME NS

**Revista del Instituto de Filosofía**

20

*Enero–Junio*
*Nº 1*
*2000*

*Universidad Central de Venezuela*
FACULTAD DE HUMANIDADES Y EDUCACIÓN

# EPISTEME NS

*semestral*





*Los números de EPISTEME NS salen alternativamente dedicados a temas de Filosofía e Historia de la Filosofía (Serie Azul) y de Lógica y Filosofía de la Ciencia (Serie Roja). Números especiales (Serie Gris).*

Suscripción anual para Venezuela: Bs. 10.000.
para el Exterior: U.S. $ 25.
Precio especial de este ejemplar: Bs. 5.500.

Favor enviar cheque pagadero a la orden de Ingresos propios de la Facultad de Humanidades y Educación de la UCV a la siguiente dirección: Instituto de Filosofía. Apartado 47342. Caracas 1041–A. Venezuela.

# EPISTEME NS

## A LOS COLABORADORES

EPISTEME NS es una revista de crítica e investigación en filosofía, abierta a todas las corrientes y estilos de pensamiento y a la reflexión en todos los ámbitos del saber filosófico, sin más requisitos que la originalidad, seriedad y rigor argumentativos. Los números salen alternativamente dedicados a temas de Filosofía, Historia de la Filosofía y Filosofía Social (Serie Azul) y de Lógica, Análisis del Lenguaje y Filosofía de las Ciencias (Serie Roja), además de las ediciones especiales (Serie Gris). La Revista, de periodicidad semestral y arbitrada según los procedimientos *ad usum*, es una publicación de circulación internacional indexada en *The Philosopher Index* y en *Revencyt*, y registrada en *Ulrich´s International Periodical Directory*, por lo que todas las colaboraciones deben ser inéditas y ser enviadas por triplicado para ser sometidas a la consideración del Comité de arbitraje. Los trabajos podrán ser de tres tipos: Artículos (no deben exceder de 10.000 palabras), Notas (5.000 palabras) y Reseñas. Todas las colaboraciones deben ser inéditas y ser enviadas por triplicado para ser sometidas a la consideración del Comité de arbitraje.

Los Artículos y las Notas y Discusiones, deben ser presentados en cuartilla tamaño carta a doble espacio con notas a pie de página (en disquete compatible con IBM en Word 6.0 para Windows o superior) y deberán ir acompañados de sendos resúmenes, en inglés y en castellano, de una extensión no mayor de diez (10) líneas, acompañados con tres palabras claves.

Las contribuciones, correspondencia, libros y revistas para recensión deben ser enviados a: Director, EPISTEME NS, UCV, apartado postal 47342, Caracas 1041–A–Venezuela.

EPISTEME NS/Revista del Instituto de Filosofía Vol. 20, No.1 (enero–junio 2000). Caracas: Ediciones de la Facultad de Humanidades y Educación, 1981.
V./ 22 cm.
2 veces al año.
Continúa: Episteme
Título de cubierta.
Director–fundador: 1981 – Juan David García Bacca
Director: 2000 – Miguel Ángel Briceño G.
ISSN: 0798–4324
1. Filosofía – Publicaciones periódicas.
Depósito Legal: pp198102DF423

Este número de la revista se publica gracias a la colaboración de la Prof. Marisa Kohn de Beker, y bajo los auspicios del Consejo de Desarrollo Científico y Humanístico de la Universidad Central de Venezuela.


Facultad de Humanidades y Educación
Universidad Central de Venezuela
Impreso en el año 2000 en
FEPUVA-UCV
Caracas–Venezuela

EPISTEME NS

Volumen 20, Nº 1, 2000

Contenido

Págs.

# PRESENTACIÓN

En 1930 Rudolf Carnap transformaba los viejos *Annalen der Philosophie* en la nueva revista *Erkenntnis*, concibiendo la operación como un "verter vino nuevo en nuevos toneles" para anunciar claramente la ruptura de contenido y método respecto de la publicación original. Salvando la distancia, concédase el mismo derecho a los menores, incluso si "a duras penas" menores. De hecho, tras dos décadas continuas de publicación, la Nueva Serie de *EPISTEME* se presenta a la comunidad filosófica "en nuevos toneles", representados por cambios importantes en la diagramación, una nueva portada y la adaptación a las disposiciones de los más importantes índices internacionales. Estamos confiados en que nuestros lectores recibirán con agrado este "new look" de la revista.

En esta misma vena de "nuevo milenio", de ajustes y remozamientos, hemos querido iniciar el año dos mil dedicando este número a un tema de extrema actualidad y de renovada reflexión filosófica, a saber: La Inteligencia Artificial y la Filosofía de la Mente. Tras un largo período de relativo estancamiento, la filosofía de la mente ha vuelto a surgir como un área excitante de búsqueda filosófica. Parte importante de este renovado interés obedece sin duda a su vinculación con la inteligencia artificial, ante los resultados asombrosos en corto tiempo obtenidos en el procesamiento computacional de la información. Influenciados durante mucho tiempo por el poder representacional del simbolismo lógico, los filósofos de la mente están ahora repensando la función de los métodos computacionales en la definición de los procesos y estados cognitivos, y considerando a la vez la pertinencia de su aplicación a la investigación en ciencias sociales. El número de

EPISTEME NS que hoy presentamos a nuestros lectores se inserta justamente en el centro de esta problemática. Cada uno de los siguientes seis ensayos refleja a su modo el compromiso del autor con una concepción más o menos "computacional" en la explicación de la mente humana. Pensamos que el resultado proporciona una visión suficientemente rica y polémica del estado actual de la cuestión en esta área de la investigación filosófica. La contribución más cuantiosa corresponde a los investigadores del Instituto de Filosofía de la Universidad Central de Venezuela y a otros involucrados en los proyectos de investigación que ahí se adelantan. Así, los artículos aquí presentados por Carlos Blank, José Burgos, Vincenzo P. Lo Monaco y Levis Zerpa M. están basados en contribuciones a simposios y congresos internacionales recientemente tenidos sobre el tema. En cambio, los ensayos del italiano Andrea Bianchi y del español Antoni Gomila Benejam fueron rigurosamente solicitados para su inclusión en este número especial. Vaya a todos ellos –especialmente a los colaboradores del exterior– nuestro más sincero agradecimiento.

El editor

EPISTEME NS, Vol. 20, Nº 1, 2000, pp.3–28

ANDREA BIANCHI

# RAPPRESENTAZIONI E ARCHITETTURA DEL MENTALE*

***Resumen:*** Según la ciencia cognitiva, las operaciones mentales son cómputos sobre representaciones. Para evitar la circularidad, se sostiene que debemos explicar cómo proceden las representaciones sin el recurso a operación mental alguna. Sugiero una estrategia diferente. Se considera que algunas operaciones son prerepresentacionales y precomputacionales, y proporcionan a la "máquina computacional" el material con el cual trabajar. En suma, propongo una arquitectura mental de varios niveles. Me apoyo en Fodor, quien explica cómo las representaciones mentales representan en términos de relaciones causales, pero desarrollo la idea de una forma más bien diferente y argumento a favor de lo que podría llamar "causalidad conceptual". La primera ocurrencia de un símbolo primitivo del lenguaje del pensamiento es causada por lo que, en razón de esto, se convierte en su significado. ¿Es esta conexión semántico–causal recuperable? Los mecanismos transductores que alimentan de representaciones a la máquina computacional "preservan la memoria" de la cadena causal desde estímulos distantes a símbolos, y fijan en consecuencia la interpretación del lenguaje formal a partir del cual queda definido el cálculo que la máquina ejecuta. Por consiguiente, la intencionalidad de lo mental reside en estos mecanismos no computacionales, casualmente subestimados por los científicos cognitivos.

**Palabras claves:** ciencia cognitiva, teoría representacional de la mente, causalidad conceptual.

**Abstract:** According to cognitive science, mental operations are computations on representations. To avoid circularity, it is held that we have to account for how it comes that representations do represent, without refer-

* Vorrei ringraziare Paolo Leonardi per i suoi commenti, e per tutto il resto.

ring back to any mental operation. I suggest a different strategy. Some operations are understood as prerepresentational and precomputational, and provide the 'computational machine' with the material on which it works. In short, I propose a mixed mental architecture. I draw on Fodor, who accounts for how mental representations represent in terms of causation, but I develop the suggestion in a rather different way, and I argue for what I would call "type-causation". The first token of a primitive symbol of the language of thought is caused by what, owing to this, becomes its meaning. Is this causal-semantic connection retrievable? The transducer mechanisms, which provide the computational machine with the representations on which it works, 'keep memory' of the causal chain from distant stimuli to symbols, and therefore fix, in a certain sense, the interpretation for the formal language on which the calculus the machine performs is defined. Hence, the intentionality of the mental rests on these not computational mechanisms, usually neglected by cognitive scientists.

**Keywords:** Cognitive science, representational theory of mind, type-causation.

Al cuore dell'approccio che negli ultimi quarant'anni si è venuto affermando, fino quasi a diventare egemonico, nello studio dei fenomeni mentali, vi è l'idea che le operazioni della mente siano *computazioni su rappresentazioni*. È importante sottolineare che questa non vuole essere una suggestione o una metafora, e neppure va confusa con l'idea, meno impegnativa, secondo cui la mente sarebbe simulabile al computer, macchina che computa rappresentazioni: se è per questo, anche l'andamento della borsa di New York può essere simulato allo stesso modo, senza che con ciò le operazioni finanziarie debbano venir considerate proprio computazioni su rappresentazioni.[1] Piuttosto, l'idea ha lo statuto di *ipotesi scientifica* – non a caso l'approccio ha preso quasi universalmente il nome di "*scienza cognitiva*" –, e come tale può essere, intesa nel suo senso letterale, vera o

[1] In altre parole, mentre la borsa di New York è un sistema che *istanzia* (vale a dire, che *è descrivibile attraverso*) una funzione computabile, la mente sarebbe, secondo l'approccio che stiamo discutendo, un sistema che ne *computa* una. Per questa spesso trascurata distinzione, vedi p.e.
Crane, T.: *The Mechanical Mind. A Philosophical Introduction to Minds, Machines and Mental Representation*, Penguin Books, London, 1995, pp. 102-104.

falsa. In particolare, se è vera, essa comporta che, realizzati in qualche modo nel cervello (dove, se no?), esistano oggetti, magari transitori, che hanno la proprietà di essere rappresentazioni, e meccanismi di un certo tipo che agiscono su di essi trasformandoli o producendone altri, e determinando, in ultimo, il comportamento del fortunato possessore del cervello in questione.

L'idea, in sé, non è affatto nuova. Nel panorama filosofico, ha fatto forse la sua prima apparizione durante il diciasettesimo secolo, per merito di Thomas Hobbes. In seguito, è stata variamente ripresa, all'interno sia della tradizione empirista che di quella razionalista. Solo nel Novecento, però, è potuta assurgere al rango di vera e propria ipotesi scientifica, attorno alla quale si è costituito un programma di ricerca che non ha ancora smesso di dare i suoi frutti. Il merito, questa volta, è stato della costruzione delle prime macchine calcolatrici, e delle riflessioni teoriche che stanno alla base delle ricerche in Intelligenza Artificiale. Il motivo è semplice. L'idea che le operazioni della mente siano computazioni vuole, al suo apparire, essere una caratterizzazione di quella cosa misteriosa che è la ragione o l'intelligenza dell'uomo: quando ragioniamo, computiamo.[2] Sfortunatamente, per quelle che erano le conoscenze del Seicento, questa caratterizzazione non poteva valere come spiegazione: l'*explanandum* sarebbe occorso surrettiziamente nell'*explanans*, ingenerando circolarità. L'intelligenza era computazione, ma la computazione richiedeva qualcosa (il famigerato *homunculus*) che facesse sì che venissero eseguite le computazioni 'giuste', quelle grazie alle quali gli esseri umani possiedono e manifestano intelligenza, e questo qualcosa, per poter svolgere adeguatamente il suo compito, sembrava dover possedere a sua volta un'intelligenza mica da poco: l'intelligenza che spiega l'intelligenza, insomma. In

---

2 "Quando un uomo ragiona, egli non fa che concepire una somma totale dall'addizione di particelle, oppure un resto dalla sottrazione di una somma da un'altra". Hobbes, T.: *Leviathan, or the Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil*, London, 1651, parte I, cap. V.

breve, l'idea era buona, ma mancavano ancora i presupposti per farne la base di un'*indagine naturalistica della mente*.[3] La svolta, come accennato, è avvenuta molto più tardi, con la costruzione delle prime macchine calcolatrici, e cioè con la palese dimostrazione che una computazione può essere completamente *meccanizzata*: il ruolo degli *homunculi* lo giocano i *programmi*.[4] La morale della storia è ovvia: se pensare è computare, e le computazioni non sono altro che particolari processi fisici, il mistero scompare. Ora, finalmente, la nozione di computazione è legittimata a figurare nell'*explanans* di una teoria scientifica che voglia render conto dei fenomeni mentali. Almeno in linea di principio, l'intelligenza (il pensiero, inteso come attività) è *naturalizzata*. A questo punto, si tratta solo di vedere se l'ipotesi tiene alla prova dei fatti, e di impostare intanto attorno ad essa la ricerca empirica, cercando di caratterizzare in dettaglio le funzioni ('cognitive') che la mente umana computerebbe e di individuare i programmi che sarebbero responsabili delle computazioni in questione. Lavoro ingrato e difficile, come sa bene chi nei decenni scorsi ha cercato effettivamente di costruire una macchina intelligente, ma che

---

3 Più degli altri, si è avvicinato a un'indagine di questo tipo David Hume, che pretendeva di aver individuato i principi (*rassomiglianza*, *contiguità*, *causa ed effetto*) che "fan sì che la mente venga trasportata da un'idea all'altra" (Hume 1739-40, libro I, sez. IV). Effettivamente, anche se manca ancora una vera e propria spiegazione meccanica del loro funzionamento, questi principi non sembrano richiedere intelligenza. Purtroppo, però, come ha rilevato Jerry Fodor, essi non sono neanche in grado di spiegarla: non possono essere quelle individuate da Hume --semplici *associazioni*-- le computazioni che sottostanno ai nostri processi mentali. Vedi Fodor, J.A., *The Modularity of Mind. An Essay on Faculty Psychology*, The MIT Press, Cambridge (Mass.), 1983, pp. 24-37; Fodor, J.A., *Fodor's Guide to Mental Representation*, "Mind", pp. 24-37; e Fodor, J.A. *A Theory of Content and Other Essays,* pp. 55-97, poi in Fodor, *A Theory of…,* cit., pp. 3-29.

4 Per i dettagli, sui quali non mi soffermerò, vedi p.e. Haugeland, J., *Artificial Intelligence. The Very Idea*, The MIT Press, Cambridge, (Mass.), 1985, cap. III-V, e Crane, *op. cit.,* cap. III. A dire il vero, già ai tempi di Hobbes Blaise Pascal aveva costruito una macchina che eseguiva somme e sottrazioni. Tuttavia, nonostante l'interesse notevole che suscitò, la sua grossolanità e la limitatezza delle operazioni che era in grado di svolgere hanno impedito che potesse essere presa a modello delle computazioni umane.

noi possiamo tranquillamente lasciare agli psicologi cognitivi.

Forse, però, abbiamo conferito un po' troppo in fretta patente di piena scientificità all'idea che stiamo discutendo. Secondo questa, come abbiamo detto, le computazioni sottostanti ai nostri processi mentali sarebbero su *rappresentazioni*. Ci sono almeno due buone ragioni per chiamare in causa la nozione di rappresentazione. La prima, nota e sulla quale non mi soffermerò, è che il comportamento (intelligente) degli esseri umani non sembra poter trovare spiegazione se non assumendo che sia determinato dall'interazione di stati *intenzionali* (e cioè, intrinsecamente rappresentazionali) come credenze e desideri: se sto digitando sulla tastiera del mio computer le parole che state leggendo, è *perché* desidero che appaiano sul monitor e credo, *inter alia*, che la pressione dei tasti che sto premendo ne determini la comparsa. Una teoria dei fenomeni mentali, allora, non può fare a meno di riconoscere l'esistenza di oggetti come le rappresentazioni.[5] Quale sia poi nello specifico il modo in cui queste giocano il loro ruolo lo chiarisce la seconda ragione, alla quale vale la pena di fare un cenno. Il fatto interessante infatti è che è la nozione stessa di computazione, del tutto indipendentemente da qualsiasi considerazione sulla natura e sul funzionamento della mente, a chiamare in causa quella di rappresentazione. Ossia, per usare uno slogan di Fodor, non ci sarebbero computazioni senza rappresentazioni ("*no computation without representation*"): un processo di qualche tipo è una computazione se e solo se opera causalmente su oggetti rispecchiando le proprietà rilevanti di altri oggetti, rappresentati o simboleggiati dai primi. Un esempio è quello delle calcolatrici tascabili, che manipolano numerali (realizzati fisicamente non importa come) in modo tale da riflettere le proprietà aritmetiche dei numeri. Così, se una teoria dei fenomeni mentali assume che le operazioni della mente siano computazioni, deve assumere anche che ciò

---

5 Per uno sviluppo rigoroso di argomenti di questo tipo si vedano Fodor, *Fodor's Guide to…, cit.,* e Fodor, *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*, The MIT Press, Cambridge (Mass.), 1987, cap. I.

che viene computato (gli oggetti su cui operano i programmi) siano rappresentazioni. L'unica alternativa che possa avere plausibilità scientifica, a quanto mi consta, è quella di vedere le operazioni della mente (cioè, quello che avviene nel cervello) come semplici processi chimico-fisici.[6] Purtroppo, si tratta di un'alternativa poco esplicativa: perché e come questi processi possono determinare il comportamento intelligente di un essere umano? La sola risposta oggi sul tappeto è in realtà quella offerta dalla scienza cognitiva: lo determinano proprio perché realizzano, di fatto, computazioni su rappresentazioni.

A questo punto, tuttavia, si ripropone la questione dello statuto scientifico della nostra idea. La nozione di rappresentazione è legittimata a figurare nell'*explanans* di una teoria dei fenomeni mentali? La domanda, in questa forma, potrebbe apparire sorprendente a chi ha seguito il discorso fin qui. Se la nozione di computazione epistemologicamente non desta più sospetti, ed essa, come abbiamo appena visto, è intimamente legata a quella di rappresentazione, come può quest'ultima creare ancora difficoltà? E, del resto, non esistono oggi forse macchine – quelle stesse macchine che hanno legittimato l'uso della nozione di computazione – che computano rappresentazioni? Naturalmente, non avrebbe alcun senso negare l'esistenza di rappresentazioni: le parole che pronunciamo, quelle scritte nei libri, i ritratti, le mappe, i cartelli stradali, e tantissime altre cose ancora, lo sono.[7] Anche le strutture di dati, sulle

---

6 Anche se spesso si definiscono computazionalisti, e anche se il livello della loro analisi è comunque più astratto, collocherei tra i sostenitori di questa alternativa anche i *connessionisti*: i processi che essi pretendono di descrivere non mi sembrano soddisfare le condizioni perché si possa parlare di computazioni.

7 Sto usando la parola "rappresentazione" in modo molto generico, a coprire tutti quei casi in cui qualcosa 'sta per', o 'raffigura', o 'rinvia a', o 'significa', o, appunto, 'rappresenta', qualcos'altro. Per dirla in termini tecnici, tutto ciò che intrattiene con qualcosa una *relazione intenzionale*. Una maggiore accuratezza -- una buona fenomenologia delle rappresentazioni -- sarebbe indispensabile, ma non ho lo spazio per occuparmene qui. In ogni caso, più avanti introdurrò la distinzione, che mi sembra importante, tra rappresentazioni e rappresentanti.

quali operano i comuni programmi per computer, senza dubbio lo sono. Il problema è invece un altro. Tutte queste cose sembrano dovere la loro rappresentatività all'intervento dell'uomo, che le produce o le sfrutta per rinviare ad altro, e che è in grado di interpretarle quando le incontra. Non sono insomma intrinsecamente rappresentazioni, anche se alcune loro proprietà, come la somiglianza nel caso dei ritratti o altre caratteristiche percettive nel caso delle parole,[8] possono favorire l'uso che ne viene fatto. D'altronde, la capacità degli uomini di usarle come tali sembra dipendere dal loro essere dotati di una mente. Così, della rappresentatività di ciò che abitualmente consideriamo una rappresentazione si deve rendere conto facendo riferimento al verificarsi di alcune operazioni mentali. E la conclusione inevitabile, a questo punto, non può che essere che la nozione di rappresentazione, così com'è, non può comparire nell'*explanans* di una teoria dei fenomeni mentali, pena, ancora una volta, la circolarità: le attività della mente spiegherebbero ciò attraverso cui si vorrebbe spiegarle.[9]

Finora, mi sono limitato a ricapitolare alcuni passaggi importanti che hanno caratterizzato la riflessione sulla mente nei decenni scorsi, senza aggiungere nulla di particolarmente nuovo. Le ultime osservazioni, più specificamente, avevano a che fare con la delicata questione della *naturalizzazione dell'intenzionalità*, oggi al centro del dibattito filosofico. Ormai quasi nessuno crede più, come facevano i pionieri dell'Intelligenza Artificiale, che la semantica scaturisca dalla

---

8 Cfr. Leonardi, P., "Come si connettono parole e cose", *Rivista di estetica,* XL, 11, pp. 17-38, 1999.

9 Detto per inciso, una circolarità simile mi sembra caratterizzare e rendere vacua la posizione di Daniel Dennett sull'intenzionalità (cfr. Dennett D.C., *Intentional Systems*, "The Journal of Philosophy", 68, pp. 87-106; poi in Dennett, *Braistorms. Philosophical Essays on Mind and Psychology,* Montgomery, Bradford Books, 1978; Dennett, *True Believers: The Intentional Strategy and Why It Works*, in Heat, 1981, pp. 53-75; poi in Denett, *The Intentional Stance*, The MIT Press, Cambridge (Mass.), [1981], 1987). Se spieghiamo (o, meglio, eliminiamo) l'intenzionalità di un sistema riconducendola ad una 'innocua' attribuzione di intenzionalità (l'"*intentional stance*" adottata nei suoi confronti), come possiamo poi rendere conto di questa attribuzione se non facendo appello all'intenzionalità *intrinseca* del sistema che la compie?

sintassi, o che il modello computazionale basti a se stesso.[10] La strategia che si sta invece tentando di seguire è quella di tener ferma l'idea che le operazioni della mente siano computazioni su rappresentazioni, e di cercare di legittimare in qualche modo l'uso della nozione di rappresentazione, sfuggendo alla circolarità. Così, si prova a partire in due la classe delle rappresentazioni: mentre quelle che incontriamo nella vita di tutti i giorni presuppongono l'intenzionalità di chi le produce o le interpreta, quelle che la mente computa sarebbero connesse *naturalmente* (causalmente o teleologicamente) al loro contenuto. Esse, cioè, rappresenterebbero ciò che rappresentano in virtù del fatto che ne sono causate (Fodor), o che la loro funzione propria (selezionata evolutivamente) è appunto quella di rappresentarlo (Ruth Millikan). Tuttavia, anche se queste proposte hanno raggiunto per merito dei loro autori un grado notevole di raffinatezza, suscitano non poche perplessità.[11] Senza entrare nel merito, quello che mi colpisce negativamente è che cause e funzioni proprie sembrano qualcosa di estrinseco, che non si riflette sul modo in cui la mente è fatta, sulla sua struttura: siamo di fronte a una metafisica del significato, più che a una sua spiegazione.[12] Proprio per questo motivo mi propongo in questo pezzo di suggerire una strategia diversa per uscire dall'*empasse*

---

10 In questo, possiamo forse vedere una vittoria 'postuma' di John Searle, che per primo, con il celebre argomento della Stanza Cinese (cfr. Searle, J., *Minds, Brains, and Programs*, "The Behavioral and Brain Sciences", 3, pp. 417-424, 1980, o Searle, *Minds, Brains and Science. The 1984 Reith Lectures*, Harvard University Press, Cambridge (Mass.), 1984), ha individuato il punto debole del computazionalismo *old style*. In ogni caso, esistono ancora delle eccezioni, almeno parziali, a quanto scritto nel testo. La più rilevante mi sembra quella della *semantica interpretazionale* proposta da Robert Cummins, che comunque non pretende di essere una soluzione al problema dell'intenzionalità. Cfr. Cummins, R..: *Meaning and Mental Representation*, The MIT Press, Cambridge (Mass.), 1989, capp. VIII-X.

11 Acero 1995 offre una chiara introduzione alle presunte soluzioni di Fodor e della Millikan, e a molte altre. Per una discussione, vedi Cummins, *cit.*, capp. V e VII, e Crane, *cit.* cap. V.

12 Un esempio di quello che intendo lo si trova in Fodor, J.A., *Concepts. Where Cognitive Science Went Wrong?* , Oxford, Clarendon Press, 1998, capp. VI e VII.

delineata nei capoversi precedenti, e di portare qualche argomento in suo favore. C'è, in fondo, un modo molto semplice per evitare la circolarità che si crea quando si sostiene contemporaneamente che le operazioni della mente siano computazioni su rappresentazioni e che la rappresentatività delle rappresentazioni sia un effetto di alcune operazioni della mente. Anziché respingere la seconda affermazione, andando alla ricerca di improbabili spiegazioni alternative, è sufficiente attenuare un po' la prima, senza con ciò rinunciare al suo ruolo esplicativo, riconoscendo che se alcune (molte) operazioni della mente sono computazioni su rappresentazioni, alcune altre potrebbero non esserlo (di più, se vogliamo metterla su un piano logico: *non possono esserlo*, se vale la seconda affermazione). La proposta, è, insomma, quella di provare a prendere in considerazione l'ipotesi di un'*architettura mista della mente*: alcune delle sue operazioni precederebbero le computazioni, e fornirebbero a queste il materiale su cui operare.[13] In questa prospettiva, oltre tutto, è forse possibile recuperare alcune delle intuizioni che sono alla base della soluzione causale di Fodor, senza essere costretti a comprare anche la sua metafisica del significato.

Come è risaputo, il quadro teorico che la scienza cognitiva è andata costruendo intorno alla sua idea di base impone di riconoscere non solo che esistano cose come le rappresentazioni mentali, ma anche che queste siano assai di più di quelle che si sarebbe ingenuamente portati a supporre. Accanto a quelle connesse agli atteggiamenti proposizionali

---

13 Anche se su questo non dirò nulla, questa precedenza andrebbe intesa non solo in senso architettonico ma anche *filo-* e *ontogenetico*. In questo modo, si attenuerebbero tra l'altro due dei principali difetti del modello computazionale: la sua scarsa plausibilità da un punto di vista evoluzionistico e il suo troppo pesante innatismo. Si potrebbe comunque obiettare che la proposta di un'architettura mista non è affatto nuova. Da tempo è noto che mentre i computazionalisti hanno ottenuto buoni risultati nella costruzione di macchine che eseguono compiti di alto livello, come il gioco degli scacchi o la dimostrazione di teoremi, in quelli di basso livello (il riconoscimento di forme, per esempio) sono i connessionisti a convincere di più. Così, molti hanno pensato che si potrebbe provare a integrare i due approcci. In ogni caso, gli argomenti di cui intendo servirmi qui sono notevolmente diversi da quelli di questo tipo.

che siamo consapevoli di avere, che costituiscono per molti il caso paradigmatico, ci sono, se una qualche versione della psicologia del profondo è corretta, quelle sottostanti a credenze e desideri inconsci perché rimossi, che pure contribuiscono alla determinazione del nostro comportamento, ma anche e soprattutto quelle, del tutto inaccessibili all'introspezione, che vengono utilizzate nei processi mentali computazionali di tipo 'inferiore', per esempio quelli legati alla produzione di enunciati sintatticamente ben formati, l'individuazione e lo studio dei quali ha giocato un ruolo di primo piano nell'affermarsi della scienza cognitiva,[14] o quelli che terminano con istruzioni all'apparato sensomotorio (si pensi a quante 'microconoscenze' sono richieste per eseguire con successo anche la più semplice delle azioni), oppure quelli che, al contrario, sarebbero responsabili dell'elaborazione dell'informazione in entrata e, in particolare –ma su questo punto esprimerò più avanti le mie riserve–, della costruzione del percetto.

Al di sotto di questa grande varietà, comunque, esiste un tratto fondamentale che deve accomunare, se l'idea che stiamo discutendo è corretta, tutte le rappresentazioni mentali. Come sappiamo, queste devono essere oggetto di computazioni, il che vuol dire che esistono meccanismi che agiscono su di esse in virtù della loro *forma*. È più che ragionevole supporre che non ogni forma sia compatibile con le operazioni di questi meccanismi, e quindi che non tutti gli oggetti, per quanto rappresentazionali, abbiano le caratteristiche per poter essere manipolati nei nostri processi mentali. In particolare, dall'analisi delle funzioni cognitive, e cioè dei compiti che i meccanismi in questione sono chiamati

---

14 Abitualmente, si fa risalire il suo atto di nascita come disciplina empirica proprio alla recensione di Noam Chomsky a *Verbal Behavior* di Skinner (Chomsky, N., *A Review of B.F. Skinner's "Verbal Behavior"*, "*Language",* 35, pp. 26-58,1959), nella quale si rileva la manifesta impossibilità di rendere conto di un fenomeno complesso come l'apprendimento del linguaggio semplicemente attraverso stimoli e rinforzi. Per una ricostruzione storica, vedi Gardner, H., *The Mind's New Science*, New York, Basic Books, 1985.

a svolgere, sembra risultare che le operazioni da eseguire siano di tipo *inferenziale* (anziché associativo, come pensava Hume), e questo impone un vincolo molto forte su ciò che le rappresentazioni mentali possono essere, perché richiede che abbiano una *struttura logica*. Il che vuol dire, sostanzialmente, che devono possedere alcune delle proprietà tipiche degli *enunciati* delle lingue naturali. Di qui alla conclusione che esista una cosa come un *linguaggio del pensiero*, nel quale sono espresse tutte le rappresentazioni mentali, il passo è breve. Anche se non proprio tutti i cognitivisti sono disposti a compierlo, io proseguirò sposando questa ipotesi, anche se ritengo che sarebbe possibile riformulare la mia argomentazione in modo che non dipenda da essa.[15]

Comunque, c'è un aspetto a proposito del quale l'analogia con le lingue naturali *non* può aiutarci. Se la significatività di queste ultime è in qualche modo spiegabile facendo ricorso alla nozione di convenzione o almeno a quella di intenzione comunicativa, lo stesso non si può dire a proposito delle rappresentazioni mentali, se non si vuole ricadere nella circolarità di cui abbiamo parlato sopra. Il problema che si pone è insomma ancora una volta quello dell'intenzionalità: che cosa conferisce a una rappresentazione mentale proprio il contenuto che ha? Grazie a cosa, per esempio, la particolare configurazione neuronale in cui è realizzata la mia credenza che domani pioverà è una rappresentazione del fatto che domani pioverà, e non del fatto che Caracas è la capitale del Venezuela?

Fare appello già qui alla nozione di causa è evidentemente senza speranza. Certamente la rappresentazione in questione è parte di un processo causale, quello che ha portato alla formazione della credenza, e che porterà, a partire da essa, ad ulteriori credenze, e a chissà cos'altro. Altrettanto certamente, però, almeno in questo caso, alla radice del processo non ci può essere ciò che la credenza

---

15 Fodor ha difeso più volte, convincentemente, l'ipotesi del linguaggio del pensiero. Vedi p.e. *Fodor, Psychosemantics. The Problem…, cit.,* app.

rappresenta. Semplicemente, ciò che la credenza rappresenta non può essere alla radice di *nessun* processo causale, per un motivo ovvio: che domani piova è solo una possibilità e, da che mondo è mondo, le possibilità – gli stati di cose possibili, alla Wittgenstein – non hanno *effetti causali* di alcun tipo (se non forse in senso metaforico, ma è bene tener presente che la nozione di causa può svolgere un qualche ruolo nella naturalizzazione dell'intenzionalità solo se è quella in uso nelle scienze naturali).[16] Se ci si rende conto che quella appena delineata non è un'eccezione tutto sommato trascurabile, ma un tratto quasi costitutivo della nostra vita mentale – ad ogni desiderio corrisponde la rappresentazione di qualcosa che non si è ancora dato, accanto alle credenze vere ci sono le credenze false, i timori sono spesso infondati, e così via – è giocoforza concludere che la relazione tra una rappresentazione mentale e ciò che questa rappresenta non è di tipo causale: non è a questo livello che ci si può servire della nozione di causa per rendere conto dei fenomeni intenzionali.[17]

A questo punto, però, ci ritroviamo esattamente nella situazione di partenza: in virtù di cosa una rappresentazione mentale rappresenta ciò che rappresenta? La strada da seguire è, credo, quella di prendere un po' più sul serio l'analogia con le lingue naturali. Se qualcosa ha una struttura logica, ha una struttura a costituenti. Se qualcosa ha una struttura a costituenti e ha significato, è probabile che debba

---

16 Non dico nulla di epistemologicamente sospetto quando affermo che ho deciso che domani non uscirò di casa perché (a causa del fatto che) pioverà. Quello che intendo è però che la mia decisione è stata determinata dal fatto che credo che domani pioverà, e non dal fatto che domani pioverà. Ciò che causa la decisione non è ciò che è rappresentato ma la rappresentazione, che a sua volta può essere causata da molte cose (dal fatto che ho visto al telegiornale le previsioni del tempo, per esempio) ma non da ciò che rappresenta.

17 Ovviamente, anche se qualche volta la terminologia adottata è stata poco felice, nessun cognitivista ha mai sostenuto una teoria causale *così* grezza. Tra l'altro, essa è incompatibile con la nostra idea: se tutte le rappresentazioni mentali fossero causate da ciò che rappresentano, le operazioni della mente non potrebbero essere caratterizzate come processi inferenziali e quindi computazionali.

il suo significato al significato dei suoi costituenti. Questo è, almeno, il caso delle lingue naturali, dove vale il principio di composizionalità. Se affermo che domani piove, la mia affermazione significa ciò che significa in virtù di ciò che significano le parole che la costituiscono, "domani" e "piove". Se lo credo senza affermarlo, e credere qualcosa è avere una rappresentazione di questo qualcosa nella 'scatola' delle credenze, e una rappresentazione è, come abbiamo detto, qualcosa di molto simile ad un enunciato, è naturale supporre che anch'essa erediti il suo contenuto dai suoi costituenti, che in ultima analisi non possono che essere i *simboli primitivi* del linguaggio del pensiero, anche se non è detto che questi simboli siano la traduzione delle parole "domani" e "piove" – la mia rappresentazione potrebbe avere una complessità molto maggiore di quella dell'enunciato italiano corrispondente. Benché io non abbia presentato quasi nessuno dei molti argomenti, in gran parte elaborati da Fodor, che si potrebbero portare a sostegno di questa ipotesi, possiamo insomma ragionevolmente supporre che valga anche per le rappresentazioni mentali quanto Ludwig Wittgenstein scriveva nel *Tractatus* a proposito dell'essenza della proposizione: "Un nome sta per una cosa, un altro nome sta per un'altra cosa ed essi sono connessi tra loro: così il tutto presenta – come un quadro plastico – lo stato di cose".[18]

In questo modo, abbiamo spostato un po' avanti quello che è il problema che stiamo cercando di affrontare. Se la rappresentatività di una rappresentazione mentale è dovuta alla rappresentatività dei simboli primitivi dai quali è costituita, a che cosa è dovuta la rappresentatività di questi ultimi? Prima di provare ad abbozzare una risposta, è opportuna una precisazione terminologica. Se i simboli primitivi 'rappresentano', nel senso molto generale per cui rinviano ad altro,[19] essi non sono però 'rappresentazioni

---

18 Wittgenstein, L.: *Logisch-philosophische Abhandlung*, "Annalen der Naturphilosophie", 14, pp. 185-262; ed. riveduta *Tractatus logico-philosophicus*, London, Kegan, 1922,1921, prop. 4.031.

19 Cfr. nota 7.

mentali', per come abbiamo usato questa nozione sin qui. Non essendo dotati di struttura logica, non possono essere oggetto di manipolazione sintattica: non sono loro gli oggetti su cui si esercita la computazione, anche se questa opera su ciò di cui sono i costituenti. Nel corso della trasformazione di una rappresentazione in un'altra in cui consiste un processo mentale del tipo in questione, essi possono venire eliminati, introdotti o spostati, ma mai modificati: sono computazionalmente *impenetrabili*, per così dire. Inoltre, *resentazioni*: giocano il ruolo che nelle lingue naturali è giocato dalle parole e non quello che è giocato dagli enunciati. Se insisto su questo, è perché mi sembra che spesso alla cosa non si presti la dovuta attenzione nell'odierna filosofia della mente, e che questo ingeneri confusioni di un certo peso.[20]

Detto questo, però, dobbiamo ancora rispondere alla nostra domanda, anche se possiamo ora riformularla diversamente. In virtù di cosa i simboli primitivi del linguaggio del pensiero stanno per ciò per cui stanno? A questo livello, forse la nozione di causa può essere di qualche aiuto. Proprio le differenze che intercorrono tra il ruolo semantico dei simboli primitivi e quello delle rappresentazioni mentali suggeriscono che potremmo essere sulla strada giusta: la critica formulata precedentemente non regge più. A differenza di una possibilità (ciò che una rappresentazione mentale rappresenta), una 'cosa' (ciò per cui un simbolo primitivo sta), può avere effetti causali. Non sappiamo ancora per quali cose i simboli primitivi propriamente stiano, per cui non sappiamo bene in che modo la causazione potrebbe realizzarsi, ma almeno alcune difficoltà di principio possono

---

20 Bertrand Russell, che insieme a Wittgenstein ha riflettuto a lungo, anche se in un contesto molto diverso, su queste questioni, scriveva che "vi sono diversi tipi di simboli, diversi tipi di relazioni tra i simboli e ciò che essi simboleggiano, e se non lo si comprende ne derivano errori gravissimi" (Russell, B.: *The Philosophy of Logical Atomism*, "The Monist", 28, pp. 495-527, e 29, pp. 32-63, 190-222, 343-80; poi in Russell, *Logic and Knowledge. Essays 1911-1950,* a cura di R.C. Marsh, London, Allen and Unwin, (1956), 1918-19, p. 185).

dirsi superate, per cui vale la pena di esplorare un po' più a fondo l'idea.

È necessario, comunque, procedere con molta cautela. Il modo più ovvio di intendere la cosa è quello per cui l'*occorrenza* di un simbolo primitivo in una rappresentazione sarebbe causata da ciò per cui il simbolo sta. Se ci si riflette sopra appena un po', però, questa ipotesi si rivela del tutto insostenibile. Supponiamo, per esempio, che nel mio linguaggio del pensiero vi sia un simbolo primitivo che sia la traduzione della parola italiana "gatto". Se ogni occorrenza del simbolo in una mia rappresentazione fosse causata da qualcosa che soddisfa il predicato "essere gatto", e se la nozione di causazione in gioco qui deve essere quella in uso nelle scienze naturali, le uniche credenze e gli unici desideri che potrei avere a proposito dei gatti sarebbero quelli che mi formo in loro presenza (la credenza che *lì* ci sia un gatto, o che *quel* gatto sia fulvo, o il desiderio che *quel* gatto si faccia accarezzare).[21] Questo è, abbastanza chiaramente, falso. A mio parere, non siamo di fronte qui solo al ben noto problema dell'*errore*, quello per cui, ad esempio in circostanze particolari di luce, posso trovarmi a credere che lì ci sia un gatto quando di fatto davanti a me c'è un cane, cosa che non succederebbe se l'oggetto davanti a me causasse l'occorrenza del simbolo che sta per esso (o, meglio, per la proprietà che esemplifica).[22] Il problema, in realtà, è molto più generale. Gli esseri umani, a differenza probabilmente da quegli animali che si trovano molto in basso nella scala dell'evoluzione, sono in grado di rivolgere la loro attenzione ad oggetti e proprietà che non sono loro immediatamente presenti – se non fosse così, non avrebbero tra l'altro sviluppato uno strumento straordinario come il linguaggio. Una conseguenza di ciò è che qui, di fronte al mio monitor, e

---

21 Questo vale solo se assumiamo che ogni relazione causale mondo-mente sia mediata dai nostri organi di senso. Di fatto, però, non mi sembra che ci siano alternative possibili, se ci si vuole muovere su un terreno non troppo lontano da quello delle scienze naturali.

22 Per il problema dell'errore, vedi Fodor, *Psychosemantics. The Problem…, cit,* cap. IV, Cummins, *op. cit.,* capp. IV-VII, e Crane, *cit.,* pp. 170-181.

in totale assenza di gatti, posso avere credenze e desideri (e, quindi, rappresentazioni) di qualsiasi tipo a proposito di essi. Se credo che nella mia stanza ora non ci siano gatti, non è ovviamente a causa del fatto che sono in presenza di gatti, e se desidero regalare un gatto a Dora, il motivo non è che ce l'ho qui davanti a me, ma che so che Dora adora i gatti e desidero renderla felice. Quest'ultimo esempio aiuta anche a trarre le conclusioni, perché è ovvio che posso intrattenere il mio desiderio di fronte a pressoché qualunque situazione esterna: nella rappresentazione soggiacente, l'occorrenza del simbolo che è la traduzione in mentalese di "gatto" non è affatto causata dal suo contenuto; piuttosto, è l'esito di un processo inferenziale, che coinvolge altre rappresentazioni. Il che vuol dire, sostanzialmente, che spesso un simbolo occorre in una rappresentazione in virtù di una *computazione*, e quindi in virtù della sua *forma* e non del suo *contenuto* (anche se, naturalmente, le computazioni devono per quanto è possibile rispettare il contenuto dei simboli sui quali operano, per cui devono soddisfare un qualche standard di *validità* inferenziale).

Scartata in questo modo l'ipotesi che sia l'occorrenza di un simbolo primitivo in una rappresentazione ad essere causata da ciò per cui il simbolo sta, quale strada possiamo ancora percorrere per trovare una risposta alla nostra domanda? La nota distinzione tipo\occorrenza (*type\token*) fornisce, credo, la chiave giusta. Se il modello computazionale che stiamo prendendo in considerazione è adeguato, ad ogni simbolo primitivo deve corrispondere una particolare configurazione neuronale.[23] Tutte le volte che il simbolo entra come costituente in una rappresentazione, la configurazione in questione si 'attiva', e sappiamo che, se quanto detto poco fa è vero, questa attivazione può essere determinata dalle cause più diverse. Perché questo avvenga, tuttavia, la configurazione deve già possedere il suo significato: quali che

[23] Su come il cervello realizzi le computazioni che ci interessano sappiamo in realtà ancora molto poco, per cui quello che scrivo non pretende ovviamente di avere alcuna attendibilità neurofisiologica.

siano i motivi per cui il simbolo compare nel corso della computazione, vi compare come rappresentante di qualcosa. Se credo che non ci siano gatti nella mia stanza, la rappresentazione che soggiace alla credenza è una rappresentazione del fatto che non ci sono gatti nella mia stanza grazie al fatto che i simboli primitivi dai quali è costituita significano ciò che significano, del tutto indipendentemente da quali siano le 'ragioni' computazionali in virtù delle quali questi simboli sono entrati a far parte della rappresentazione. Il quadro che si delinea, insomma, è quello di una configurazione neuronale che viene attivata nelle circostanze più diverse, pur essendo la realizzazione fisica di un simbolo con un significato ben specifico. Le possibilità, a questo punto, sono due: o la configurazione è connessa a ciò cui è connessa 'per sua natura', e cioè intrinsecamente, oppure vi è un qualche momento nella vita mentale in cui essa assume il significato che ha. La prima alternativa è puramente metafisica, e conduce verso un innatismo radicale che non consiste affatto in una spiegazione o naturalizzazione dell'intenzionalità ma piuttosto in una sua ipostatizzazione.[24] Se vogliamo prendere sul serio il problema, invece, è la seconda possibilità che merita di essere esplorata. Il modo per farlo è, credo, quello di pensare che sia la prima attivazione della configurazione ad essere in qualche modo causata da quello che diverrebbe, in seguito a ciò, il suo significato. Anche se le successive occorrenze del simbolo possono, come abbiamo visto, essere determinate dalle cose più diverse, esso mantiene il suo significato, perché le occorrenze sono

---

[24] Naturalmente, questa osservazione non ha come bersaglio polemico l'innatismo in quanto tale. La scienza cognitiva è legata a doppio filo con l'idea che almeno una parte dei meccanismi (i 'programmi') che determinano le computazioni che la mente esegue sia specificata geneticamente, e quindi innata. Questo *innatismo dei meccanismi* è però cosa molto diversa dall'*innatismo dei contenuti*, secondo il quale possederemmo innatamente non solo i programmi, ma anche parte dei 'dati' su cui questi operano. Anche Fodor, che pure a suo tempo ne è stato grande sostenitore, ha forse preso ultimamente un po' le distanze da questa posizione (cfr. Fodor, *Concepts. Where Cognitive…, cit.,* cap. VI; vedi anche Margolis, E., "How to Acquire a Concept", *Mind and Language,* XIII, 3, pp. 347-369, 1998).

appunto occorrenze *di quel simbolo-tipo*.[25] Se vogliamo, il quadro che risulta da queste considerazioni presenta alcune assonanze con quello delineato da Saul Kripke nella seconda lezione di *Naming and Necessity* a proposito dei nomi propri delle lingue naturali: come per questi ultimi, anche per i simboli primitivi del linguaggio del pensiero andrebbe tenuto ben distinto il momento della loro introduzione da quello del successivo 'uso', che spesso avviene in assenza di ciò di cui il simbolo è simbolo.[26]

La proposta, comunque, va incontro a molte difficoltà, che proprio queste assonanze aiutano a mettere in luce. Gli usi di un nome proprio presuppongono una 'recuperabilità' della relazione intenzionale: almeno in linea di principio, la cosiddetta 'catena causale' deve poter essere risalita, perché solo così è possibile discriminare un uso corretto da un uso scorretto del nome. Il problema è che quest'ultimo non porta su di sé informazioni su ciò che lo ha originato. La soluzione, nel caso delle lingue naturali, è semplice: c'è almeno qualcuno, colui che ha compiuto il battesimo, che *sa che* il nome è il nome di *quello* specifico individuo, ed è a questa conoscenza che noi in ultima analisi deferiamo, quando usiamo il nome con "l'intenzione di usarlo con lo stesso riferimento". Prima di andare a vedere la forma che il problema assume nel caso dei simboli primitivi del linguaggio del pensiero, osserviamo subito che per esso una soluzione di questo tipo non sarebbe accettabile, perché farebbe leva sull'intenzionalità (il *sapere che*) della mente, che è proprio

---

25 Non mi è chiaro se questo è quello che ha in mente Fodor quando parla di *dipendenza asimmetrica*. In ogni caso, mentre lui sembra assumere questa come *la* spiegazione, io ritengo si tratti di qualcosa che deve essere ulteriormente spiegato. Per una critica forse simile, a cui non segue però la parte propositiva, vedi Crane 1995, pp. 180-181.

26 Cfr. Kripke 1972. Mentre quando il nome viene introdotto l'individuo al quale comincia a riferirsi è di solito presente ("normalmente, colui che battezza è in qualche modo in un rapporto di conoscenza diretta con l'oggetto che denomina", p. 96, n.), nei casi di successivo uso tutto ciò che è semanticamente rilevante è l'"intenzione di usare lo stesso riferimento": "quando il nome 'viene trasmesso da un anello all'altro', il ricevente del nome deve […] aver l'intenzione di usarlo con lo stesso riferimento di colui dal quale lo ha appreso" p. 96.

ciò che stiamo invece cercando di spiegare.

Di fatto, per quanto riguarda quello che ho chiamato "problema della recuperabilità", un simbolo primitivo del linguaggio del pensiero si trova in una situazione assai difficile. I motivi, mi sembra, sono almeno due. Il primo è legato al fatto che le rappresentazioni devono avere, come abbiamo detto, formato linguistico, e, tipicamente, un simbolo di una lingua (a differenza di una combinazione di simboli) non reca alcuna informazione su ciò di cui è simbolo. Vale a dire: se è una 'cosa' specifica ad aver causato la prima attivazione di una configurazione neuronale, divenendo così ciò per cui questa sta, non c'è però nulla nella configurazione che la connetta intrinsecamente ad essa. È *quella* configurazione, ma avrebbe potuto benissimo essere un'*altra*, e da un semplice esame di essa non è possibile risalire alla sua causa (confronta: "Andrea" è il mio nome, ma avrebbe potuto benissimo essere "Giorgio", e l'ispezione della parola "Andrea", per quanto accurata, non permette di stabilire che io ne sia il portatore). A questo argomento, comunque, è possibile replicare sostenendo che l'analogia con la lingua naturale sarebbe in questo caso fuorviante, in quanto potrebbe essere proprio la spiegazione in termini causali della significatività di un simbolo primitivo del linguaggio del pensiero a differenziarlo da una parola dell'italiano e ad offrire la chiave della soluzione al problema della recuperabilità: una configurazione neuronale, forse, è in grado di 'tenere memoria' di ciò che l'ha una volta attivata, e di 'riconoscerlo' quando si ripresenta; almeno potenzialmente, contiene tutta l'informazione necessaria. Anche se fosse così, però, e questo a me sembra l'argomento decisivo, se la mente fosse solo una macchina computazionale –se tutte le sue operazioni consistessero in computazioni su rappresentazioni –queste informazioni le sarebbero *completamente inaccessibili*, e il problema della recuperabilità rimarrebbe irrisolto. Il punto è semplicemente che, per quanto materialmente complesso possa essere, un simbolo primitivo è, quasi per definizione,

*computazionalmente semplice* (o, come abbiamo già detto, *impenetrabile*). I meccanismi che operano sulle rappresentazioni nelle quali entra come costituente sono sì in grado di riconoscerlo in virtù della sua forma, ma non di scomporlo o di analizzarlo, e quindi, di estrarre da esso – dalla sua struttura fisica– l'informazione che (forse) contiene. Le possibilità, a questo punto, sono due: o non tutte le operazioni della mente sono computazioni su rappresentazioni, oppure il nesso causale attraverso cui si spiega la significatività dei simboli primitivi non è recuperabile.

Di fronte a ciò, il computazionalista potrebbe compiere un estremo tentativo di difesa, negando la rilevanza stessa del problema della recuperabilità. In fondo, potrebbe sostenere, una volta che il simbolo è 'causato', e ha acquisito quindi il suo contenuto, il resto viene da sé. La macchina computazionale avrebbe solo bisogno di essere avviata in questo modo, dopo di che, dati i simboli primitivi, alcune rappresentazioni base e degli algoritmi efficienti, i processi mentali andrebbero come dovrebbero andare: le computa-zioni, come è noto, operano sulle rappresentazioni solo in virtù della loro forma, senza alcun bisogno di 'guardare' al loro contenuto. Anche se la sintassi non crea la semantica, come pensavano i pionieri dell'Intelligenza Artificiale, perlomeno la rispetta, e ciò sarebbe più che sufficiente.

Questa difesa, però, non tiene. Il problema è che, per il buon funzionamento della macchina uomo, il nesso semantico-causale *deve* poter essere recuperato. Supponiamo, ancora una volta, che nel mio linguaggio del pensiero vi sia un simbolo primitivo che sia la traduzione della parola italiana "gatto". Se le cose stanno come le ho presentate, questo simbolo è stato causato (e cioè, la configurazione neuronale che lo realizza è stata per la prima volta attivata) da un qualche gatto – presumibilmente, il primo che ho visto, o il primo al quale ho rivolto la mia attenzione. Ovviamente, questa causazione deve essere stata mediata dagli organi di senso,[27] che hanno raccolto informazione (i bastoncelli della

[27] Cfr. nota 21.

retina sono stati urtati da particolari fotoni, per esempio) e l'hanno trasferita al cervello mediante il sistema nervoso. Questo processo di trasmissione è puramente fisico: non dobbiamo pensare a una computazione, ma semmai a qualcosa di vagamente simile all'impressionamento di una lastra fotografica. È solo nel cervello, dunque, che l'informazione assume formato simbolico, con l'attivazione di quella configurazione neuronale che funzionalmente svolge il ruolo dell'equivalente della parola italiana "gatto". A questo punto, la 'macchina', secondo la posizione del computazionalista estremo che stiamo discutendo, potrebbe 'dimenticare' il processo, senza che ciò costituisca un problema: quanto segue sono solo manipolazioni sintattiche, che l'evoluzione ha in qualche modo reso affidabili. Questo, tuttavia, non può che essere falso. Una capacità che certamente tutti noi abbiamo, e di cui una teoria dei fenomeni mentali deve rendere conto, è quella di *riconoscere* un gatto, perlomeno quando lo incontriamo in circostanze ambientali normali.[28] Questo riconoscimento, ovviamente, è alla radice della fissazione di molte credenze ("Ecco un gatto!"), le cui rappresentazioni sottostanti conterranno il nostro simbolo primitivo. Ma, se le cose stanno così, dobbiamo ammettere che in qualche modo la mente ha riconosciuto l'informazione che gli organi di senso le hanno trasmesso come l'informazione che era intervenuta nel processo causale che aveva originato il simbolo: essa 'ricorda' il processo, e cioè è in grado di recuperare il nesso semantico.

Naturalmente, si potrebbe sempre sostenere che non esista nel linguaggio del pensiero un simbolo primitivo che sia la traduzione della parola italiana "gatto". In questo caso, il riconoscimento che avviene potrebbe essere l'effetto di un processo puramente inferenziale, l'esito di una serie di

---

[28] In un contesto solo in parte diverso, quello della semantica (cognitiva) delle lingue naturali, Marconi 1997 ha rilevato, in questo stesso spirito, come le posizioni che riducono la *competenza lessicale* a una competenza puramente *inferenziale* non hanno gli strumenti per affrontare il problema del riconoscimento. Secondo Marconi, alla componente inferenziale della competenza si dovrebbe accostare una componente *referenziale*.

computazioni su rappresentazioni di livello più basso, i cui costituenti ultimi sono connessi causalmente a proprietà 'più sensoriali' di quella di essere gatto. Così, però, il problema viene soltanto spostato. L'informazione che gli organi di senso trasmettono in presenza di un gatto giunge comunque al cervello in formato non simbolico, e deve essere reidentificata, per dare il via al processo inferenziale che conduce al riconoscimento dell'animale: le configurazioni neuronali che realizzano i simboli primitivi delle rappresentazioni su cui opera la computazione devono attivarsi, e devono attivarsi in virtù del fatto che l'informazione che arriva viene riconosciuta come quella *rilevante* – quella che ha causato la loro prima attivazione.

In qualche misura, gli scienziati cognitivi sono consapevoli di tutto ciò. Quello che non può essere seriamente messo in discussione, almeno, è che, mentre una macchina computazionale non può che avere come ingressi e come uscite rappresentazioni, e queste rappresentazioni, come abbiamo visto, devono esibire un particolare formato, un essere umano si trova a interagire con un ambiente esterno, dal quale estrae informazione attraverso gli organi di senso. Queste informazioni (le stimolazioni prossimali) non sono rappresentazioni, nel nostro senso; vanno invece considerate come meri segnali fisici. Perché la macchina computazionale possa accedervi, essi devono essere trasformati in rappresentazioni del formato giusto, che possano essere 'lette' e manipolate sintatticamente. A svolgere questo lavoro di trasformazione sono i cosiddetti "*trasduttori*". Non si tratta, ovviamente, di meccanismi che eseguono computazioni, quanto piuttosto di "sistemi analogici che convertono le stimolazioni prossimali in segnali neurali covarianti in modo più o meno preciso", segnali che possiamo vedere come i simboli primitivi del nostro linguaggio del pensiero.[29]

---

[29] Vedi Fodor, *Fodor's Guide to…,* cit., cap. II, in cui si afferma, tra l'altro, che "è difficile pensare che un calcolatore possa far a meno di esibire dei meccanismi di trasduzione, se appena è interfacciato con il mondo esterno", p. 42.

Tuttavia, l'importanza di questi meccanismi di trasduzione è stata, a mio parere, gravemente sottovalutata. Se ci si riflette un po' su, è infatti possibile vedere proprio in essi la soluzione al problema che stiamo affrontando. Se guardiamo al lavoro che svolgono, dobbiamo concludere che, in qualche modo, questi meccanismi contengono il *codice* che associa ogni simbolo primitivo alla sua causa, e quindi, se il percorso che abbiamo seguito sin qui è corretto, al suo contenuto. Per dirla un po' metaforicamente, essi giocano per il linguaggio del pensiero il ruolo che per i nomi propri della lingua naturale è giocato dal *sapere che* di colui che ha introdotto il nome: conservano l'informazione che consente di 'riproiettare' un simbolo su ciò di cui è simbolo.[30] Se vediamo nella mente, come è implicito nell'idea secondo cui le sue operazioni sarebbero computazioni su rappresentazioni, un *calcolo definito su un linguaggio formale*, i meccanismi di trasduzione fissano, tarskianamente, la *realizzazione* di questo linguaggio: su di essi riposa, in ultima analisi, la sua *intenzionalità*.

Siamo, così, giunti praticamente al termine della nostra storia. Anche se ho ripreso l'idea fodoriana di rendere conto della rappresentatività delle rappresentazioni mentali in termini di causazione, l'ho sviluppata in un modo abbastanza differente, andando a vedere come l'architettura della macchina uomo potrebbe essere compatibile con essa. Se le cose stanno come le ho presentate, alcune operazioni, *prerappresentazionali* e *precomputazionali* (ma che, tra un attimo vedremo perché, vanno a pieno titolo qualificate come "mentali"), spi omparire nell'*explanans* di una teoria dei fenomeni mentali.

Perché una proposta di questo tipo, tutto sommato

---

30 La metafora non va comunque presa troppo sul serio. Non avrebbe ovviamente senso dire che i meccanismi *sanno che* quel simbolo sta per quella tal cosa, o che grazie ad essi lo sa la mente. Assai più modestamente, essi si limitano ad associare, in modo affidabile, alla stessa causa lo stesso effetto. Questa soluzione del problema della recuperabilità non va, insomma, affatto a sostenere una visione cartesiana del mentale, per cui alla mente sarebbero trasparenti le sue operazioni e i suoi contenuti.

abbastanza naturale, e che ha per certi versi una lunga storia (si pensi al Descartes scienziato, o ai meccanicisti francesi del Settecento), non è ancora affiorata, a quanto ne so, nel corso del dibattito sulla naturalizzazione dell'intenzionalità che ha coinvolto, negli ultimi anni, chi si è occupato dei fondamenti della scienza cognitiva? La mia impressione è che la causa sia una certa 'foga' computazionalista, che ancora ci si porta dietro dai tempi d'oro dell'Intelligenza Artificiale, in cui sembrava che tutto fosse facile e il trionfo dietro l'angolo. Allora, di meccanismi di trasduzione non si parlava affatto. Anche oggi, malgrado si riconosca la loro necessità, lo si fa di malavoglia, e si cerca di sminuirne il più possibile la portata. Fodor, per esempio, si esprime così: "il modo più naturale di interpretare gli output dei trasduttori consiste nella specificazione della distribuzione degli stimoli sulla 'superficie' dell'organismo".[31] A questo punto, entrerebbero in gioco gli analizzatori di input, "sistemi esecutori di inferenze", le quali "hanno come 'premesse' delle rappresentazioni trasdotte delle configurazioni degli stimoli prossimali, e come 'conclusioni' le rappresentazioni delle caratteristiche e della distribuzione degli oggetti distali".[32] La costruzione del percetto, dunque, sarebbe di fatto un processo computazionale. Di più: il significato dei simboli primitivi del linguaggio del pensiero, quelli che entrerebbero nelle rappresentazioni che fungerebbero da premesse di questo processo, sarebbe tale per cui sarebbe impensabile far ricadere su di essi il peso dell'intenzionalità del mentale. Di qui, la ricerca di altre strade, che però, come ho avuto occasione di dire, hanno poco della spiegazione e molto della speculazione metafisica. Dal momento tuttavia che né una necessità logica né un dato empirico vietano di pensare che il lavoro svolto dai meccanismi di trasduzione sia molto più ampio – le ricerche di gestaltisti, gibsoniani e connessionisti, in modi diversi, vanno in questa direzione – l'alternativa che propongo, quella di un'architettura mista della mente, mi

31 Fodor, *The Modulary of…,* cit., p. 42.
32 *Ibid.*

sembra praticabile. Se la costruzione del percetto fosse l'opera non della macchina computazionale ma di complessi processi fisici di elaborazione dell'informazione,[33] e solo al termine di questi processi facessero la loro comparsa simboli e rappresentazioni, molti misteri, a partire da quello dell'intenzionalità, scomparirebbero. La cosa, come ho già accennato, sembra tra l'altro plausibile anche da un punto di vista ontogenetico e, soprattutto, filogenetico, perché nella scala dell'evoluzione i sistemi in grado di percepire – o almeno, ma a me sembra poi la stessa cosa, di estrarre informazione dall'ambiente circostante – hanno preceduto, di milioni di anni, quelli dotati di linguaggio.[34]

Università del Piemonte Orientale e Università di Bologna

[33] Se i connessionisti hanno ragione, questi processi fisici istanziano una funzione computabile, e quindi sono simulabili al calcolatore. Questo non vuol dire, ovviamente, che siano processi computazionali. Cfr. nota 1.

[34] Per un'indagine sul rapporto tra percezione, intenzionalità e linguaggio che va nella direzione che ho cercato di suggerire, vedi Leonardi 1999.

## RIFERIMENTI BIBLIOGRAFICI


Acero, J.J.: *Teorías del contenido mental*, in Broncano, pp.175-206, 1995.

Broncano ,F.: (ed.), *La mente humana*, Madrid, Editorial Trotta, 1995.

Chomsky, N.: *A Review of B.F. Skinner's "Verbal Behavior"*, "Language", 35, pp. 26-58, 1959.

Davidson, D. - Harman G. (ed.): *The Semantics of Natural Language*, Dordrecht, Kluver, 1972.

Dennett, D.C.: *Intentional Systems*, "The Journal of Philosophy", 68, pp. 87-106; poi in Dennett [1978], 1971.

Dennett, D.C.: *Braistorms. Philosophical Essays on Mind and Psychology*, Montgomery, Bradford Books, 1978.

Dennett, D.C.: *True Believers: The Intentional Strategy and Why It Works*, in Heat [1981], pp. 53-75; poi in Denett, 1987.

Dennett, D.C.: *The Intentional Stance*, Cambridge (Mass.),The MIT Press, [1981], 1987.

Fodor, J.A.: *The Modularity of Mind. An Essay on Faculty Psychology*, Cambridge (Mass.), The MIT Press, 1983.

Fodor, J.A.: *Fodor's Guide to Mental Representation*, "Mind", pp. 55-97; poi in Fodor [1990], pp. 3-29, 1985.

Fodor, J.A.: *A Theory of Content and Other Essays*, Cambridge (Mass.),The MIT Press, 1990.

Fodor, J.A.: *Concepts. Where Cognitive Science Went Wrong*, Oxford, Clarendon Press, 1998.

Gardner, H.: *The Mind's New Science*, New York, Basic Books, 1985.

Heat, A.F. (ed.): *Scientific Explanation: Papers Based on Herbert Spencer Lectures Given In the University of Oxford*, Oxford, Clarendon Press, 1981.

Hume, D.: *A Treatise of Human Nature*, London, 1739–40.

Kripke, S.: *Naming and Necessity*, in Davidson - Harman [1972], pp. 253-355 e 763-769; Oxford, Blackwell, 1980, 1972.

Marconi, D.: *Lexical Competence*, Cambridge (Mass.), The MIT Press, 1997.

Margolis, E.: *How to Acquire a Concept*, "Mind and Language", XIII, 3, pp. 347-369, 1998.

Russell, B.: *Logic and Knowledge. Essays 1911-1950*, a cura di R.C. Marsh, London, Allen and Unwin, 1956.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 29–49

CARLOS BLANK

# PENROSE Y LA INTELIGENCIA ARTIFICIAL

**Resumen:** Podemos señalar el año de 1950, cuando Alan Turing publicó en la revista *Mind*, su artículo "Maquinaria computacional e Inteligencia", como el punto de partida de la IA. Desde entonces, las ideas desarrolladas en este artículo han sido objeto de las más variadas discusiones, a veces en contra y otras a su favor. En 1979 fue publicado el libro de Hofstadter, *Gödel, Escher, Bach*, libro que habría de convertirse en un gran éxito editorial así como en una de las mejores defensas jamás escritas de la IA. Diez años más tarde, apareció el libro de Penrose, *La nueva mente del emperador,* un libro que también habría de convertirse en un gran éxito editorial así como en uno de los ataques mejor escritos contra la IA hasta el momento. Nuestro artículo se centra en las ideas desarrolladas por Penrose en contra de la IA, ideas que pueden ser consideradas, en buena parte, como una versión de las ideas que ya antes habían sido planteadas por el importante filósofo norteamericano, John Searle.

**Palabras claves:** Filosofía de la mente, inteligencia artificial, máquina computacional.

**Abstract:** We may consider the year of 1950, when Alan Turing published, in *Mind*, his article "Computing Machinery and Intelligence", as the starting point of IA. Since then, the ideas taken up in this article had been subject of a wide discussion, sometimes in their favour and sometimes against them. In 1979 was published Hofstadter's, *Gödel Escher, Bach,* one of the best defenses of IA ever written, that became an editorial success. Ten years later appeared Penrose's ***The Emperor's New Mind***, a book that also became an editorial success and was considered as the best attack against IA up to that moment. Our article is focused on the ideas taken up by Penrose against the IA, ideas that are in some extend a corrected version of the ideas previously stated by the important American philosopher, John Searle.

**Keywords:** Philosophy of mind, artificial intelligence, computational machine.

## *INTRODUCCIÓN: Técnica y civilización.*

La capacidad que tiene el hombre de crear herramientas se remonta a los albores de la civilización y, aun más allá, al propio proceso de la evolución humana. En gran medida estas herramientas son las responsables de esta evolución más que su resultado, aunque también es cierto que entre el hombre y sus herramientas se produce una interacción permanente. El hombre crea un mundo humano en la misma medida en que crea un mundo técnico, así como un mundo de normas y ritos. Cocinar los alimentos, enterrar los muertos, prohibir el incesto, utilizar herramientas son, como el lenguaje, costumbres o instituciones que delatan la presencia de la conciencia humana y permiten crear un espacio propiamente humano frente a lo puramente natural.

A menudo se considera a la ciencia y a la técnica como las responsables de la alienación y deshumanización del hombre. Se piensa que una visión mecanicista de la naturaleza y del hombre constituye la fuente de todos los males y que ha despojado al hombre de su verdadera naturaleza humana racional. Nada más alejado de la verdad, pues ha sido precisamente el conocimiento de los mecanismos subyacentes en la naturaleza lo que ha permitido un mayor dominio del hombre sobre ella, así como también el que haya podido irse liberando, hasta cierto punto, de las fuerzas ciegas que gobiernan el mundo de la naturaleza, haciendo posible la creación de una segunda naturaleza para así liberarse de tareas pesadas y repetitivas que antes él solía hacer. Como señalaba Bertrand Russell en una oportunidad, los inventos de la era moderna han hecho más por la liberación de la mujer que cualquier movimiento feminista.

Somos más humanos a medida en que es posible liberarnos de las barreras naturales. No podemos volar, pero entonces inventamos aparatos con los cuales podemos hacerlo. No podemos respirar bajo el agua como los peces, pero también

inventamos dispositivos que nos permiten realizar tal hazaña. Y así sucesivamente. El mundo técnico constituye una compleja red de "palancas" y "poleas" que multiplican las limitadas fuerzas naturales del hombre, incluyendo la de los animales. Como en el viejo mito prometeico, ha sido la debilidad natural del hombre la que le ha permitido justamente ganarle terreno a la naturaleza, hacer de la necesidad virtud, de la debilidad su fuerza.

Pero el hombre no sólo ha creado máquinas que pueden impulsar su fuerza mas allá de sus fronteras naturales o aprovechar las propias fuerzas de la naturaleza, sino que ha creado también herramientas y medios que economizan y potencian considerablemente su fuerza mental o psíquica. Desde la escritura cuneiforme, pasando por la imprenta, hasta llegar a los modernos chips y discos láser, el hombre se ha visto en la necesidad de ingeniar medios para el almacenamiento, transmisión y conservación de información. Asimismo ha inventado reglas de cálculo, así como las herramientas que les faciliten la realización de tales cálculos. Dentro de esta evolución podemos mencionar la propia utilización de piedras -de donde deriva el cálculo su nombre-, la invención del 0, los ábacos, calculadoras rudimentarias, hasta llegar a las modernas computadoras y el lenguaje binario que las alimenta. Si las primeras máquinas multiplican y ahorran energía física, las segundas hacen lo propio con la energía mental.

Como en todo proceso de cambio importante, se generan al comienzo resistencias. La introducción de máquinas en el proceso productivo generó bastantes recelos, pues se pensaba que estas máquinas, al desplazar la fuerza de trabajo humano, restaban oportunidades a los que trabajaban con los medios artesanales tradicionales. Lo que entonces no podía verse era el potencial de nuevas oportunidades que la introducción de la máquina generaba. Algo similar ha ocurrido con la introducción de las modernas computadoras. A lo que hay que añadir un elemento adicional. Si el hombre no se siente disminuido o minusvalorado por la presencia de máquinas que potencian su fuerza física, que son capaces de realizar tareas

que el hombre es incapaz de realizar físicamente, la competencia en el plano mental es otra cosa. La razón de ello es que suponemos que este plano mental es una prerrogativa típicamente humana, lo que eleva al hombre por encima de la naturaleza y lo hace específicamente humano.

> Hace tiempo que nos hemos acostumbrado a la maquinaria que nos supera ampliamente en las tareas *físicas.* Esto no nos causa desasosiego. Antes bien nos complace tener aparatos que nos llevan por tierra normalmente a grandes velocidades –más de cinco veces más rápido que el más veloz atleta humano- o que puedan cavar hoyos o demoler estructuras que nos estorban a velocidades que dejarían en ridículo a equipos compuestos por docenas de hombres. Aún estamos más encantados de tener máquinas que nos permiten hacer físicamente cosas que nunca antes habíamos podido hacer; pueden llevarnos a los cielos y depositarnos al otro lado del océano en cuestión de horas. Tales logros de su parte no hieren nuestro orgullo. Pero el poder *pensar*; eso sí que ha sido siempre una prerrogativa humana. Después de todo, ha sido esa capacidad física la que, al traducirse en términos físicos, nos ha permitido trascender nuestras limitaciones físicas y la que parecía ponernos encima de nuestras criaturas hermanas. Si las máquinas pudieran llegar a superarnos algún día en esa cualidad importante en la que nos habíamos creído superiores, ¿no tendríamos entonces que ceder esa superioridad a nuestras creaciones?[1]

El texto de Penrose que acabamos de citar resume brillantemente todo lo anterior. Como él señala, la posibilidad de que las máquinas puedan pensar no es algo totalmente nuevo, sólo que la moderna tecnología le ha dado a esta cuestión un nuevo impulso. La idea de que existan máquinas que puedan realizar tareas que hasta ahora creíamos exclusivamente humanas, nos hace vernos menos exclusivos de lo que nos creíamos. Todo aquello que nos hace exclusivos no es tan exclusivo desde el momento en que una máquina es capaz también de realizarlo. De eso se trata precisamente cuando hablamos de Inteligencia Artificial, de saber si es posible que una máquina realice todas las tareas intelectuales que antes creíamos prerrogativa exclusivamente humana. Uno de los pioneros de la IA y de la ingeniería de las computadoras, el

---

1 Penrose, R., *La nueva mente del emperador*, Barcelona, Grijalbo–Mondadori, 1995, p. 23 y ss.

matemático inglés Alan Turing, denominaba la posición ya descrita como la objeción "del avestruz" ("the 'heads in the sand' objection), pues se trata más de una objeción, o de un prejuicio, que de un argumento, como otros que él enumeraba. Para él esta objeción se reduce a que "nos gusta creer que el hombre es de alguna manera sutil, superior al resto de la creación. Aún es mejor si puede demostrarse que ha de ser *necesariamente* superior, porque entonces no hay peligro de que pierda su posición de autoridad."[2]

Como se desprende de todo lo dicho hasta ahora, la polémica en torno a la IA tiene connotaciones muy particulares para el ser humano, toca fibras bastantes profundas de nuestro ser, por lo que incluso resulta difícil verla como una cuestión de mero diseño tecnológico posible, como algo en torno a lo cual podemos ser indiferentes. Quizás la cuestión de fondo resida en la necesidad de definir lo que nos hace verdaderamente humanos. Desde este punto de vista, el mayor aporte de esta discusión consistirá en dar un nuevo paso en la tarea, posiblemente infinita, de conocernos a nosotros mismos.

A continuación analizaremos algunas de las implicaciones de este tema y finalmente desarrollaremos la posición asumida por Roger Penrose sobre el particular.

## *PLANTEAMIENTO GENERAL: ¿Pueden pensar las máquinas?*

La pregunta de si pueden pensar las máquinas nos conduce inevitablemente a que aclaremos primero lo que entendemos por *pensar.* Aunque ello pudiera parecernos obvio, no lo es, ni mucho menos. De lo que entendamos por *pensar* dependerá la respuesta que podamos dar a esta pregunta. La actividad genérica de pensar incluye una gran variedad de tareas específicas, muchas de ellas diferentes entre sí, aunque podamos agruparlas a todas ellas bajo esta denominación.

Tomemos, por ejemplo, la definición que dio de "pensar" Descartes, quien hizo de esta actividad, ni más ni menos, el punto de partida de todo su sistema y método. Para él una

---

2 Turing, A., "¿Piensan las máquinas?", en Newman, J.R. (Ed.,), *El mundo de las matemáticas,* Barcelona, Ediciones Grijalbo, 1983, p. 46.

cosa que piensa significa "algo que duda, que concibe, que afirma, que niega, que quiere, que no quiere, que también imagina, y que siente."[3] Para ser la definición de un gran matemático, lo menos que podemos decir es que se trata de una definición bastante vaga. Si la tomásemos como definición operacional de la inteligencia humana, diría, al mismo tiempo, muy poco o demasiado. Es evidente que la facultad de pensar no es para Descartes lo mismo que inteligencia. O si lo es, debe incluir en ella aspectos como la capacidad de percibir e imaginar, enriqueciendo nuestra concepción de inteligencia. Lo curioso es que Descartes negaba estos atributos a los animales, considerándolos meras máquinas. Aunque pueda parecer una visión muy cruel de los animales, resulta una visión bastante real de lo que debe ser una máquina. En realidad, no es descabellado afirmar que en estos aspectos un animal puede ser bastante superior a cualquier máquina inteligente. Para poner sólo un ejemplo, la capacidad *espontánea* que tiene un perro de reconocer el rostro o el tono de voz de su amo –no digamos olfatearlo- es algo que excede la capacidad de cualquier computador digital de última generación. Por lo demás, el atribuirle cierto grado de inteligencia a un animal no es nada descabellado, tomando en cuenta que son nuestros parientes lejanos, si hemos de hacerle caso a Darwin. La posición de Descartes puede estar totalmente descarriada en relación con los animales, pero no en lo que debe ser una máquina. Si a alguna máquina le pudiésemos atribuir algún grado de conciencia o, más aún, de autoconciencia, dejaría de ser una máquina.

Uno de los argumentos que se esgrimen con mayor frecuencia en contra de las pretensiones de la IA es que ninguna de las operaciones que es capaz de realizar un computador está acompañada de conciencia. Como señalaba acertadamente Turing, gran parte de los argumentos en contra del supuesto carácter pensante de las computadoras puede ser considerado como una variante de este argumento básico. Por

---

[3] Descartes, R. "Meditations", en *Oeuvres et Lettres,* París, Gallimard, 1953, p. 278.

esta misma razón debemos revisarlo detenidamente, más aun tomando en cuenta que será uno de los argumentos básicos que Penrose va a utilizar en contra de la IA.

Como en todo argumento, lo importante no es tanto el contenido como la forma de que está revestido. A menudo suele considerarse a la conciencia como algo específica y exclusivamente humano, de tal modo que cualquier cosa que no sea humana carece de conciencia. Seguramente así pensaba Descartes cuando consideraba a los animales iguales a máquinas, así como nuestro cuerpo. Pero este argumento resulta viciado desde su origen. La aparente contundencia de este argumento reside en su total circularidad. Es evidente que si definimos a una máquina como algo propiamente no humano y consideramos a la conciencia como un rasgo exclusivamente humano, entonces es inevitable concluir que las máquinas son incapaces de tener conciencia. Por otro lado, si encontrásemos una máquina capaz de ser consciente, a partir de ese momento dejaría de ser tal, sería también humana. Más que una imposibilidad real, estaríamos en presencia de una imposibilidad lógica, de una contradicción, de la negación de una tautología. Estamos simplemente en presencia de una definición: Toda máquina es, por principio, incapaz de ser consciente. Como en el ejemplo clásico: "Todos los cuervos son negros" se reduce a afirmar que: "Todos los cuervos negros son negros". Este argumento es irrefutable, pero carece por eso mismo de contenido alguno. Si el problema que estamos analizando no es algo puramente formal, sino algo también material, que debemos decidir de forma empírica, ateniéndonos al veredicto final de la experiencia, el argumento anterior resulta totalmente insuficiente, al menos tal y como lo hemos planteado hasta ahora. Debemos, entonces, revisar nuestra premisa y dejar abierta la posibilidad de que existan cosas conscientes no-humanas, que exista la posibilidad de encontrar "cuervos no negros", de encontrar "máquinas conscientes". Debemos mantener esta actitud abierta, pues de lo contrario nuestro argumento puede ser simplemente el reflejo de un prejuicio, traduce un orgullo o una vanidad

inmerecida: solamente nosotros podemos ser realmente conscientes, nada –o nadie- más. Admitir, en fin, que una máquina es consciente, nos conduciría reconocer que no es la conciencia lo que nos hace humanos, como pensaba Descartes.

Un argumento bastante socorrido contra la IA, formalmente similar al anterior, consiste en excluir de la esfera del pensar todo aquello que hasta ahora es capaz de hacer un computador y considerar como pensamiento solamente aquellas cosas que no ha sido capaz de realizar. Lo que caracterizaría propiamente el pensamiento humano sería el conjunto de actividades que no ha sido o no es capaz de ejecutar. Es evidente que siempre podemos considerar cosas que una computadora nunca ha realizado aún o que, tal vez, nunca podrá realizar: como enamorarse, disfrutar de un apetitoso bocado o sentir un orgasmo[4]. Pero no es esto el objetivo de la IA, sus pretensiones se reducen al modelaje de las actividades intelectuales de la mente humana –aunque en las actividades antes mencionadas es posible que exista un ingrediente intelectual. Como indicaba Turing, la debilidad de este argumento es la de todo argumento inductivo: se cree que una computadora es capaz de realizar solamente lo que ha hecho hasta ahora. Cuando apenas se empezaban a desarrollar las primeras computadoras había una gran cantidad de cosas que se pensaba que eran imposibles de realizar por ellas y que hoy son algo absolutamente rutinario[5]. Podríamos contraargumentar, también inductivamente, que como gran parte de lo que antes estaba fuera del alcance de las computadoras ya no lo está, entonces nada impide que aquellas cosas que hasta ahora no ha podido realizar estén bajo su alcance en el futuro. Aunque tampoco es un argumento concluyente, pues nin-

---

4 Puede verse una lista bastante completa de estas incapacidades en Turing, op. cit., p. 48 y ss.

5 Puede encontrarse una lista bastante completa de estas actividades en Hofstatder: *Gödel, Escher, Bach,* New York, Vintage Books, 1980, p. 601 y ss. Hofstatder denomina este argumento como el "Teorema de Tesler", que formula así: "once some mental function is programmed, people cease to consider it as an essential ingredient of 'real thinking'. The ineluctable core of intelligence is always in that next thing which hasn't yet been programmed" o, más brevemente, "AI is whatever hasn't been done yet". *Ibid.* p.601.

gún argumento inductivo lo es, lo cierto es que la IA surge "when mechanical devices took over any tasks previously performable only by human minds."[6] A juzgar por todos los avances de los últimos años, los avances en redes neuronales y sistemas de procesamiento en paralelo por ejemplo, el desafío de la IA, de que toda actividad mental humana es realizable también, en principio, por una máquina, no debe ser considerado descabellado ni ser tomado a la ligera.

Otro argumento que también aparece con frecuencia contra la IA, consiste en señalar la imposibilidad de que algo puramente material pueda dar origen a algo inmaterial como la conciencia. Dejando a un lado las posibles complicaciones a las que nos conduciría el clasificar a algo como material o inmaterial, este argumento tampoco resiste un análisis serio. La conciencia, al igual que la vida, tiene su origen en la organización de la materia inorgánica, por lo que no es imposible, en principio, reproducir las condiciones iniciales que dieron origen a ambas. Por improbable que ello haya sido, es innegable que han sido procesos materiales y físicos los que han dado origen a la vida primero y luego a la conciencia. Por más emergentes y novedosas que sean la vida y la conciencia, ellas son el resultado de complejos procesos físicos y químicos que apenas comenzamos a comprender. El reciente descubrimiento del genoma humano es un ejemplo, así como toda la investigación que se lleva a cabo sobre la neurofisiología del cerebro humano. Y si nos ponemos a comparar, la posibilidad de crear inteligencia por medios artificiales ha resultado algo mucho más sencillo que crear vida de manera totalmente artificial. El que la conciencia tenga una base material no dice nada a favor o en contra de la IA. Y, como veremos más adelante, tampoco el hecho de que el cerebro sea, como sostiene Searle, un "wetware" y no un "hardware".

De nuevo, todo el asunto nos remite a dilucidar qué entendemos por pensar. Si definiésemos la actividad de pensar solamente como la actividad de resolver problemas, indepen-

6 *Ibid.*, p.601.

dientemente de que en este proceso interviniese o no lo que llamamos conciencia, resulta difícil negarle esa virtud a las computadoras. Como siempre, esto nos resuelve un problema pero nos crea otro, pues ahora debemos aclarar nuestro concepto de "resolver problemas". La resolución de problemas abarca un dominio bastante amplio de operaciones posibles. Existe una parcela de problemas que pueden resolverse de manera mecánica mediante la aplicación de una determinada regla de cálculo, conformada por una serie de pasos finitos previamente definidos. De hecho, las modernas computadoras surgen, en parte, del diseño de máquinas, como la máquina de sumar y restar de Pascal o la de multiplicar y dividir de Leibniz, que realizaban operaciones relativamente complejas en un tiempo bastante menor al que le llevaría a un ser humano promedio. Si definiésemos a las computadoras como meros dispositivos de cálculo, tendríamos que admitir no solamente que son inteligentes, sino que son mucho más inteligentes que cualquiera de nosotros o que cualquier genio humano en esa área. (Hay personas que muestran una gran eficiencia en esta área, aunque en otras pueden ser considerados bastante atrasadas. Los franceses han acuñado el término de "idiot savant" para referirse a estos casos. La película "Rain Man", donde se describe un caso particular de autismo, también ejemplifica lo anterior.)

Si pusiéramos a competir al más eficiente de los calculistas humanos con una computadora moderna el resultado sería evidente. En gran parte la eficiencia de una máquina reside precisamente en que solamente es eso: una máquina, en que está diseñada para realizar procesos puramente mecánicos y mecanizables, en los cuales no interviene para nada la conciencia. La rapidez con que una máquina realiza cálculos complejos no deja ninguna duda de que se trata de eso: una máquina. No sería difícil distinguir, al respecto, entre las respuestas de una máquina y un ser humano. Pero qué ocurriría si existiese una máquina capaz de responder a todas nuestras preguntas como lo haría un ser humano – lo cual supone cierta homogeneidad entre las respuestas que daría

cualquier ser humano-, si sus respuestas fueran indistinguibles de las que daría un ser humano ante el mismo problema o situación. ¿Deberíamos atribuirle pensamiento en ese caso? De contestar afirmativamente a esta cuestión, y sólo en ese caso, el problema se reduciría a: ¿es posible, en principio, construir una máquina de acuerdo a la especificación anterior? Todo ello nos lleva a otra cuestión básica: ¿Podemos considerar a todo proceso mental como el producto de algún proceso mecánico subyacente, de alguna regla algorítmica y computable de procedimiento? Si la respuesta es sí, entonces nada impide que una máquina pueda, en principio, reproducir dicho procedimiento y en esa misma medida atribuírsele una mente y en consecuencia, pensamiento. Como veremos más adelante, buena parte del argumento de Penrose contra la IA consiste en la imposibilidad de establecer semejante homología.

Suele señalarse también que una computadora es capaz de realizar actividades propias del ser humano, pero no es capaz de integrar todas estas funciones dentro de un todo unificado, sólo puede trabajar serialmente y tomando en cuenta cada función a la vez. Este argumento desconoce todo el avance que se ha hecho en el campo de las redes neuronales y en el intento de modelar el funcionamiento en paralelo del cerebro. Se trata de un campo abierto de investigación, sobre el cual no se ha establecido la última palabra.

Una de las objeciones más interesantes ha sido la señalada por filósofos del lenguaje como Searle. Para él, el funcionamiento inteligente de una computadora se mueve en el plano de la mera simulación. Sus respuestas pueden ser indistinguibles de las de un ser humano, pero ello no se traduce en ningún momento en una comprensión real del mundo. La razón de ello es muy simple: el lenguaje que ellas utilizan se mueve en un ámbito puramente sintáctico, están diseñadas para funcionar con un lenguaje formal, carente de significado.

Otra de las objeciones que suelen plantearse en contra de la IA –denominada por Turing como "la objeción de Lady Lovelace" dirigida a la "máquina analítica" de Babbage– con-

siste en señalar que las máquinas sólo hacen lo que le pedimos que hagan, son meras herramientas al servicio del hombre. Incluso en los casos en que es capaz de autoprogramarse o programar a otras, aun así requiere de la inteligencia humana que la ha diseñado con tal fin. Este argumento, si lo podemos considerar como tal, ha sido utilizado, desde hace tiempo, en diversos contextos y responde a la necesidad de encontrar una causa primera, la causa de toda causa o el principio de toda la serie de causas y efectos, evitando así un indeseable regreso al infinito.[7] En otras palabras, responde a la necesidad de encontrar un fundamento para la existencia de una serie causal. Pero este argumento se revierte contra quienes lo utilizan, pues también cabe preguntarse por qué hemos de detenernos en el hombre como el artífice de las máquinas. ¿No podemos considerar también a los hombres como máquinas programadas a partir de un código genético que es trasmitido de generación en generación? ¿No somos los seres humanos producto también de las "programaciones" que recibimos a través de la educación y del medio que nos rodea? ¿Hasta qué punto nuestras pautas de pensamiento son también el producto de "programas" innatos o adquiridos? ¿Nuestros instintos y nuestras conductas inteligentes no son también el reflejo de estas programaciones previas? ¿No formamos acaso parte del mismo universo y debemos obedecer a las mismas leyes? ¿Estamos, en realidad, en una situación tan diferente de las máquinas que usamos?

Todo esto nos lleva a plantearnos cómo es la estructura básica del universo y las leyes que lo rigen. ¿Vivimos en un mundo gobernado por una necesidad ciega, en un mundo regido por leyes estrictamente deterministas? Y en ese caso ¿cómo podemos ser agentes libres, cómo podemos tomar decisiones de modo voluntario y libre? Surge entonces la pregunta: ¿qué tan libre somos realmente? ¿No estamos permanentemente sometidos a serias restricciones en nuestra forma

---

7 En relación con este argumento, así como a otros que hemos discutido aquí, puede consultarse a Marx Wartofsky, *Introducción a la filosofía de la ciencia,* Madrid, Alianza Editorial, 1976, Tomo II, p. 487 y ss.

de pensar y de actuar? ¿Podemos incluso definir la libertad fuera de este marco restrictivo? ¿No serán, entonces, la conciencia y la libertad humanas meras ilusiones, que surgen por nuestra ignorancia de los mecanismos subyacentes que los originan? ¿Podrá una máquina diseñada de acuerdo a las especificaciones de la física cuántica, una máquina probabilística, borrar las diferencias entre el pensar humano y el de una máquina? ¿Estará ahí la clave para el desarrollo de la IA en el futuro?

De este modo, la pregunta inicial, aparentemente simple, sobre si las máquinas pueden pensar, nos ha remitido a toda una serie de problemas relacionados entre sí, para desembocar en el planteamiento acerca de nuestro papel dentro del inmenso universo del que formamos parte ¿Somos meros engranajes dentro de un mecanismo majestuoso que apenas comenzamos a entender? ¿Supone nuestra presencia alguna diferencia importante en el marco del universo, responde a alguna finalidad específica? En caso de que así sea: ¿cuál pudiera ser ésta? ¿O somos simplemente el producto de un mero azar sin ningún propósito? A continuación analizaremos las respuestas que da Penrose a todos estos problemas.

## *LA NUEVA MENTE DEL EMPERADOR*

Este es el sugestivo título de la obra de Roger Penrose. Se trata de una obra solamente comparable a la que diez años antes escribiera Hofstadter.[8] Si esta última constituye una de las mejores defensas de la IA, la primera, como señala Martín Gardner en el prefacio, "es el ataque más poderoso nunca escrito contra la IA". Aunque cada una de estas obras defiende tesis opuestas, ambas son similares en algunos aspectos. Ambos son libros realizados por especialistas en su campo, Hofstadter en el área de la computación y Penrose en el de la física, aunque van dirigidos a un público mucho más amplio que el del especialista y han resultado ser todos unos "best sellers". En la medida de lo posible, se trata de evitar un len-

8 *Vid supra,* notas 1 y 5.

guaje demasiado técnico y cuando no hay más remedio que introducirlo, realizan una labor bastante pedagógica de aclaración conceptual A pesar de la gran variedad de temas que son tocados en ambos libros, podríamos decir que cada uno propone una clave básica para su comprensión. Si el de Hofstadter está escrito siguiendo el espíritu lúdico de Lewis Carroll, el de Penrose puede ser condensado en el espíritu lúdico de Christian Andersen, del niño que ve lo que ningún adulto quiere ver y se atreve a decirlo: El emperador está desnudo. Podría objetársele a Penrose que si la cuestión es tan evidente, y todo el problema de la IA puede ser respondido sin problemas por un niño, por qué nos ha sometido a un recorrido tan extenso y se ocupa de cuestiones abstrusas de lógica, matemática o física. Pero ello no hace sino indicar lo complejo que puede ser el llegar a una respuesta simple, lo difícil que puede resultar para un adulto conservar la visión infantil, el asombro metafísico del niño ante los enigmas que nos plantea la realidad, así como la capacidad de ver lo obvio que los demás no son capaces de ver. Recobrar esa visión, al mismo tiempo simple y profunda de un niño, constituye, para Penrose, la clave para la solución de nuestro problema. Aunque nuestra pregunta inicial pueda parecer simple y pueda requerir una respuesta simple, encierra, como ya hemos visto, "temas profundos de la filosofía".

"¿Qué significa pensar o sentir? ¿Qué es una mente? ¿Existen realmente las mentes? Suponiendo que sí existen, ¿en qué medida dependen de las estructuras físicas a las que están asociadas? ¿Podrían existir mentes independientemente de tales estructuras? ¿O son simplemente los modos apropiados de funcionar de (ciertos tipos apropiados de) estructuras físicas? En cualquier caso, ¿es necesario que las estructuras relevantes sean de naturaleza biológica (cerebros) o podrían también estar asociados con componentes electrónicos? ¿Están sujetas a las leyes de la física? ¿Cuáles son, de hecho, las leyes de la física?"[9]

Pero si la clave para la solución de nuestro problema es

9 Penrose, *op. cit*, p. 24.

relativamente sencilla, debemos desentrañarla siguiendo ciertos pasos previos en los cuales surgen estas cuestiones.[10] La mayoría de estas cuestiones ya han sido asomadas en nuestro planteamiento general del problema, aunque tendremos que detenernos en ellas con más detalle con la finalidad de analizar la posición asumida por Penrose al respecto.

## *PENROSE Y LA IA*

Uno de los pioneros de la IA es, sin duda, el matemático inglés, Alan Turing. No sólo sus planteamientos, sino también las posibles objeciones a estos planteamientos –como ya tuvimos oportunidad de destacar-, fueron claramente formulados en un artículo publicado originalmente en la revista *Mind,* en 1950, bajo el título de "Computing Machinery and Intelligence". Aunque ya antes se había ocupado de cuestiones de computabilidad, cálculo efectivo o recursividad, es de este ensayo que arranca el planteamiento central de la IA. Allí Turing reformula la pregunta, ¿pueden pensar las máquinas?, en términos de un experimento imaginario, de un test o, como él lo llama, un "juego de imitación". Los integrantes de esta prueba son inicialmente: Un hombre (A), una mujer (B), y un interrogador (C), que puede ser hombre o mujer. La prueba consiste en que C pueda saber cuál de los dos es el hombre y cuál la mujer. Para ello debe permanecer en una habitación separada de A y B, desde donde debe de plantear

---

10 Como ya mencionamos, el libro de Penrose abarca una gama muy amplia de temas. Temas de metalógica, como la máquina de Turing o el teorema de Gödel, que plantean cuestiones interesantes sobre computabilidad o recursividad. Cuestiones matemáticas como el origen de los números complejos y su relación con el conjunto de Mandelbrot en la geometría fractal; las telesaciones y los cuasicristales. Así como también las diversas nociones de espacio que se maneja en física: el espacio de fases de la mecánica clásica, el espacio tensorial métrico de Riemann en la mecánica relativista o el espacio de Hilbert de la mecánica cuántica, y las diversas hipótesis cosmológicas que se discuten actualmente. En muchos de estos campos, como el de las telesaciones o el de la cosmología relativista, Penrose ha sido un investigador original. Obviamente no podemos seguir todos los pasos previos que sigue el propio autor. Y aunque su libro constituye una unidad, el tema de la IA es tratado principalmente en el primer y último capítulo, por lo que pueden ser leídos con relativa independencia de los otros capítulos más técnicos.

una serie de preguntas a ambos que puedan arrojarle pistas de quién es quién. Para que el resultado dependa exclusivamente de las respuestas de ambos y no de otros factores como la voz, las respuestas deben ser pasadas por escrito por medio de un dispositivo apropiado. A esto hay que añadir que si A debe dar respuestas engañosas o pistas falsas, B en cambio debe ayudar al interrogador a encontrar la respuesta correcta. Pues bien, la pregunta acerca de si una máquina puede pensar o no se reduce para Turing a sustituir en el caso anterior A por una máquina. O dicho de otro modo: si las respuestas de A son iguales a las de B para un tercero C, entonces podemos afirmar que A piensa. Todo se reduce, así, a la posibilidad de diseñar un programa y una máquina apropiada que pueda superar con éxito esta prueba. Se dice, entonces, que si una máquina es capaz de superar con éxito "el test de Turing" ello es una prueba de que piensa. [11]

Para Penrose este test puede ser considerado "como aproximadamente válido en su contexto". Desde un punto de vista puramente operacional, tendría validez, pues "el operacionalista diría que el computador *piensa* con tal de que *actúe* de manera indistinguible de cómo lo hace una persona cuando está pensando."[12] El test de Turing es perfectamente compatible con este marco operacional, con un modelo conductista o de caja negra, para el que son irrelevantes los procesos que se dan internamente; lo importante son las preguntas de entrada y las respuestas de salida, los inputs y los outputs. Por cierto que éste es el modelo a partir del cual se construyen las computadoras. En cuanto a considerarlo como un modelo apropiado para explicar la mente o el comportamiento humano, existen dudas razonables.

La cuestión de fondo que subyace a este planteamiento es la de que toda actividad, ya sea la de una máquina o la de un ser humano, traduce la ejecución de una regla mecánica de decisión, es expresión de algún algoritmo. Es evidente que muchas actividades humanas y procesos de pensamiento son

---

11 Cf. Turing, *op. cit.*, pp. 36-60.
12 Penrose, *op. cit.*, p. 27.

altamente mecanizables. El diseño de cualquier máquina responde a esta posibilidad y representa una gran utilidad en la medida en que nos libera de tareas que son puramente mecánicas y repetitivas. Como ya lo señalaba Whitehead, el avance de la civilización puede medirse por la cantidad de actividades que pueden realizarse de manera mecánica, sin pensar en ellas, dejando disponible el pensamiento para otras tareas. Pero ¿podemos afirmar válidamente que todas las actividades mentales humanas son mecanizables, pueden ser ejecutadas por una máquina? Como señala Penrose, existe cierta ironía en el hecho de que sea el propio Turing el que nos dé la pista para responder negativamente a esta pregunta. La razón, sin entrar en complicadas cuestiones lógicas, es la siguiente: sabemos que una máquina ejecuta un algoritmo si es capaz de detenerse una vez que obtiene un resultado, pues éste es definido precisamente como un proceso de pasos finito que da un resultado determinado correcto. Pero el mismo Turing llegó a la conclusión de que no es posible diseñar una máquina universal que determine para toda máquina si va a detenerse o no. Que una máquina se detenga o no es algo que no puede resolverse o decidirse de forma mecánica. Ni siquiera el campo de las matemáticas se deja encerrar, todo él, en la ejecución de un algoritmo y eso que constituye su mayor fuente y origen. Como repite de diversas formas Penrose, "la decisión sobre la validez de un algoritmo ¡*no* es ella misma un proceso algorítmico!"[13] y "la verdad matemática *no* es algo que averigüemos simplemente utilizando un algoritmo."[14] Para él existe un elemento esencialmente no algorítmico en el pensamiento humano consciente y que, por lo tanto, no puede ser reproducible o modelable por una máquina, no importa lo compleja que esta pueda ser. El pensamiento humano tiene un ingrediente no-algorítmico, que no se deja reducir a un proceso mecánico repetible.

La posición de Penrose comparte algunas de las ideas se-

13 *Ibid.*, p. 514.
14 *Ibid.*, p. 518.

ñaladas anteriormente por Searle, aunque difiere en algunos aspectos de él. Para Searle, la posibilidad de que una máquina ejecute satisfactoriamente un algoritmo no es ninguna prueba de que piensa. Un computador digital está diseñado para manipular signos, para operar dentro de un nivel exclusivamente sintáctico, para relacionar un signo con otro, y carece de esa dimensión semántica, más aún pragmática, del lenguaje humano, de la que se surge la comprensión y la conciencia[15].

Al respecto Penrose dice: "El punto importante de Searle –y pienso que tiene bastante fuerza- es que la mera ejecución de un algoritmo correcto *no* implica en sí mismo que haya tenido ninguna comprensión."[16] Pero aunque se trata de un argumento bastante fuerte, no lo considera concluyente. En particular, considera que Searle expresa una confusión generalizada sobre el tema, y concede demasiado a la IA, cuando señala que el cerebro humano, como, en principio, cualquier cosa, puede ser considerado un computador digital. El libro de él se propone precisamente "demostrar por qué, y quizá cómo, esto *no* tiene que ser así."[17]

El otro punto en que su posición difiere de la de Searle es el de la importancia que este último le confiere al material del que están hechos los cerebros humanos, en comparación con el material con el cual construimos un computador. Para él, éste no es ningún aspecto relevante y "en sí mismo esto no me parece señalar el camino hacia una teoría de la mente científicamente útil."[18] Refiriéndose a posibles argumentos contra la IA, también añade que "el simple hecho de que el computador pudiera estar construido a base de transistores, cables y similares en lugar de neuronas, venas, etc., *no* es, propiamente dicho, el tipo de cosas que consideraría evidencia en co-

---

15 Hemos desarrollado más ampliamente la posición del importante filósofo del lenguaje en "Searle y la IA", en *Análisis*, Unimet, 2000. Allí analizamos el experimento de la "habitación china" propuesto por Searle y que Penrose analiza con bastante detalle.

16 Penrose, *op. cit.*, p. 43.

17 *Ibid.*, p. 48.

18 *Ibidem.*

ntra."[19] En este aspecto particular, su posición se aparta de Searle y se aproxima, de modo curioso, a la tesis de la IA, contra la cual va dirigido todo su ataque, pues comparte con ella la tesis de la indiferencia con relación a la ubicación del algoritmo: en la mente humana, en la máquina o en un mundo platónico de verdades matemáticas, como lo hace Penrose al final.

Para nuestro autor, el carácter no-algorítmico de la conciencia tiene mucho que ver con la propia creatividad y originalidad del pensamiento humano. Contrariamente a aquellos que ven en el inconsciente la fuente de esta creatividad y originalidad, él considera que son los procesos inconscientes los que se realizan de modo algorítmico.[20] La originalidad depende más del rechazo consciente a una idea que de la ocurrencia o propuesta de origen inconsciente.[21] Los procesos conscientes sobre los que se basan en el rechazo o no de alguna idea, aparece en la emergencia de las intuiciones instantáneas, valorizaciones estéticas o de visualizaciones no-verbales, pues la idea de que el pensamiento consciente debe estar acompañado de un lenguaje verbal es posiblemente la expresión de un juicio o prejuicio filosófico.[22] Todo esto es relevante en este contexto, pues a Penrose le "parece poco concebible que la verdadera inteligencia pudiera estar presente a menos que estuviera acompañada de la consciencia."[23] Él reconoce que este es el tema que realmente le preocupa y le importa: ¿qué papel desempeña la conciencia en el vasto universo del cual emerge?

Podemos resumir todo este argumento del siguiente modo: Si la verdadera inteligencia debe estar acompañada de consciencia y si la consciencia es irreducible a un proceso de

---

19 *Ibid.*, p. 32.
20 Cf. *Ibid.*, p. 510 y *passim.*
21 Cf. *Ibid.*, p. 524.
22 Cf. *Ibid.*, pp. 518-27, *passim.* Véase también nuestros trabajos: "La inserción de los valores en el contexto de la racionalidad científica", en *Revista venezolana de filosofía,* Ucab, 1989, y "Modelos y metáforas: La función de la analogía en la investigación científica", en *Análisis,* Unimet, 2000.
23 Penrose, *op. cit.*, p. 505.

tipo recursivo o algorítmico, entonces parece inevitable reconocer que la tesis de la IA es errada. La idea de que un computador puede modelar la conciencia o, más aún, la autoconciencia humana, a través de un proceso de referencia o autoreferencia, es una idea completamente superficial de lo que todo ello representa. Aunque no lo mencione en este contexto, seguramente está pensando en Hofstadter cuando dice: "Pero un programa de ordenador que contenga dentro de sí (digamos como subrutina) alguna descripción de otro programa de ordenador no hace al primer programa consciente del segundo; ni ningún aspecto *auto*-referencial de un programa le hace *auto*–consciente."[24]

La conciencia humana y el humano pensar son algo demasiado precioso para dejarlos en manos de las máquinas o para que ellas decidan, por nosotros, lo que es el pensar humano. Siempre he pensado, al analizar este tema, que la real fuente de preocupación acerca de la posibilidad de que las máquinas piensen realmente, deberá surgir solamente cuando ellas sean capaces de plantearse y comprender esa misma pregunta en relación consigo mismas y con nosotros, pues allí estaría el germen de la conciencia humana. Y es posible que pudiesen llegar a la conclusión, tomando en cuenta muchas de las cosas que el hombre ha sido capaz de realizar, que realmente no pensamos. Quizás éste sería el tipo de experimento imaginario que se plantearía un niño para destacar lo que a él le resulta obvio: ¡El Emperador está desnudo!, lo cual significa que las computadoras carecen de inteligencia real, pues carecen de la comprensión y de la conciencia que acompaña a cualquier ser humano. Esta visión de niño no depende de todos los posibles tecnicismos, argumentos o de la construcción de complejos modelos teóricos, como los que solemos emplear los adultos, a menudo sólo para inflar nuestros propios egos. Se trata más bien de una captación intuitiva de lo que resulta evidente, pues "por encima de estos tecnicismos está el sentimiento de que es realmente 'obvio' que la mente *consciente* no puede trabajar como un compu-

---

24 *Ibid.*, p. 508.

tador, incluso aunque mucho de lo que está realmente implicado en la actividad mental podría hacerlo."[25]

F.A.C.E.S.
Universidad Central de Venezuela.

[25] *Ibid.*, p. 555. Esto no impide, en cualquier caso, que podamos seguir considerando la tesis de la IA como bastante interesante y útil, incluso fascinante, en la medida en que nos acerca a una mayor compresión de nosotros mismos y de nuestro lugar en el universo.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 51–61

JOSÉ E. BURGOS

# SIMULANDO UN ASPECTO DEL PROBLEMA MENTE–CUERPO EN SISTEMAS NEURALES ARTIFICIALES[1]

**Resumen:** El problema mente–cuerpo consiste de un número de aspectos relacionados entre sí, uno de los cuales tiene que hacer con nuestra habilidad para clasificar ciertos tipos de relaciones entrada–salida (E-S) que observamos en nosotros mismos y en otros. Esta habilidad da lugar a la búsqueda de explicaciones bajo diversas condiciones ambientales y es necesaria (aunque no suficiente) para resolver el problema, en la medida en que la solución requiera de una clasificación teóricamente efectiva de relaciones E-S. Por tanto, cualquier factor que afecte dicha habilidad, afectará, en la misma dirección, nuestra posibilidad de resolver el problema. Se Muestra que un factor que afecta la habilidad de un sistema neural artificial para clasificar las relaciones E-S de otro sistema neural artificial del mismo tipo es la complejidad relativa de ambos sistemas. Específicamente, un elemento neural del tipo McCulloch–Pitts (MP) no puede clasificar totalmente y las relaciones E-S de otro elemento MC de igual o mayor complejidad.

**Palabras claves:** Red neural, inteligencia artificial, conexionismo.

**Abstract:** The problem mind-body consists on a number of aspects related to each other. One of has to do with our ability to classify certain types of relations of input - output that we observe in ourselves and in the others. This ability permits the search of explanations in different environmental conditions. This ability is necessary (though not adequate enough) to solve this problem, as long as the solution requires an effective theoretical classification of the relations input-output. Therefore, any factor that affects this

[1] Trabajo presentado en el V Congreso Nacional de Filosofía y I Coloquio Nacional de Lógica. Caracas, 20 al 23 de noviembre de 1999.

ability would affect in the same way our possibility of solving the problem. In this way is shown that a factor that affects the ability of an artificial neural system to classify the relations input – output of another artificial neural system of the same kind, is the same type of relative complexity of both systems. A neural element of the type McCulloch Pitts (MP) can not fully classify the relations input – output of another MC element of the same or bigger complexity.

**Keywords:** Neural networks, artificial intelligence, connectionism.

A pesar de haber ocupado una porción sustancial de la historia de la filosofía y la ciencia, el problema mente-cuerpo permanece como un tópico de intenso debate donde aún no se avizoran acuerdos conceptuales ni metodológicos. Las discusiones han sido numerosas y muy diversas, pero una tendencia relativamente común ha sido considerar el problema como algo característicamente humano, en al menos dos sentidos. Primero, si sólo los seres humanos poseemos mente, entonces sólo los humanos podemos tener algo como el problema mente-cuerpo, aun cuando los seres no humanos puedan tener lo que llamaríamos 'problemas corporales', en la medida en que posean cuerpos materiales. Segundo, si sólo los seres humanos podemos pensar, reflexionar, o discutir filosófica y científicamente acerca de la relación entre mentes y cuerpos, entonces solamente los humanos podemos tener algo como el problema mente-cuerpo, aun cuando concedamos que los seres no humanos poseen mente.

En este segundo sentido (el cual parece más plausible para la filosofía contemporánea de la mente), los seres no humanos no pueden tener el problema mente-cuerpo, en la medida en que represente un problema filosófico-científico. Los seres no humanos pueden tener otros problemas, tales como procurarse alimento, asegurarse un territorio, o procrear (problemas corporales típicos). Los seres humanos, por nuestra parte, enfrentamos problemas similares, pero sólo nosotros pareciéramos ser capaces de tener algo como el problema mente-cuerpo. Posiciones particulares sobre el problema, pues, surgen como propuestas acerca de cómo podemos resolverlo (las cuales derivan en el dualismo, el materia-

lismo, el funcionalismo y el conductismo) o como la hipótesis de que no podemos resolverlo (la cual deriva en el naturalismo trascendental). Para bien o para mal, sólo los seres humanos podemos tener y solucionar (o, por lo menos, intentar solucionar) el problema. En este sentido, las discusiones sobre el problema mente-cuerpo tienden a ser fuertemente antropocéntricas.

Ciertamente, la idea de que los seres no humanos puedan tener algo como el problema mente-cuerpo parece poco plausible. Después de todo, nadie ha visto (que yo sepa) gorilas o chimpancés (mucho menos murciélagos) arguyendo entre sí acerca de la naturaleza de la consciencia, la naturaleza del cerebro, o de cómo lo segundo da lugar a lo primero. Lo mismo se aplica, *mutatis mutandis*, a sistemas tales como redes neurales artificiales y máquinas de Turing. No obstante, lejos de ser monolítico, el problema es un compuesto de múltiples aspectos relacionados entre sí, algunos de los cuales representan problemas más específicos que no son exclusivamente humanos. En esta medida, intentar reconstruir tales aspectos en referencia a ciertos sistemas no humanos puede ser provechoso para nuestras discusiones del problema como un todo.

El presente trabajo constituye un intento de ese tipo. Específicamente, me intereso en el problema de clasificar relaciones ambiente-conducta, estímulo-respuesta, o entrada-salida (E-S, en adelante) en cierto tipo de sistema artificial, a saber, el elemento procesador neural introducido por McCulloch y Pitts (1943), al cual me referiré de ahora en adelante como 'MP'. En su forma más básica, la interpretación que propongo consiste en el siguiente escenario virtual. Sean $MP_K$ y $MP_O$ dos elementos del tipo MP, tales que $MP_K$ representa una especie de sujeto cognoscente que observa a $MP_O$ y $MP_O$ representa una especie de objeto siendo observado por $MP_K$. La pregunta que planteo es la siguiente: ¿Qué condiciones debe satisfacer $MP_K$ para que sea capaz de clasificar las relaciones E-S de $MP_O$? Mi objetivo principal es dar respuesta a esta pregunta, para luego explorar algunas de sus implica-

ciones para la filosofía de la mente en general. Tal exploración involucrará un grado considerable de generalización, en al menos dos sentidos. Primero, como ya lo he mencionado, la clasificación de relaciones E-S observadas en otros y en nosotros mismos constituye sólo un aspecto del problema mente-cuerpo, de tal manera que la exploración en cuestión exige determinar la forma en que aspecto y problema están ligados entre sí. Segundo, los sistemas neurales artificiales parecen conceptualmente pertinentes al problema mente-cuerpo, en la medida en que constituyan abstracciones de la estructura y función de sistemas nerviosos naturales, y en la medida en que supongamos que éstos juegan un papel central en la determinación o mediación de fenómenos mentales. Sin embargo, no sólo los sistemas neurales artificiales son construcciones teóricas simplificadoras, lo cual plantea el problema de su relevancia descriptiva y explicativa respecto a sistemas nerviosos naturales, en especial el de los humanos, sino que también un énfasis sobre estos sistemas nos obliga, de entrada, a tomar un camino materialista o, por lo menos, conexionista en la filosofía de la mente, que no es lo que aquí pretendo. Antes de pasar a la interpretación en cuestión, pues, permítaseme elaborar brevemente el primer sentido, ligando aspecto y problema de tal manera que se minimice el tipo de sesgo filosófico al cual obliga el segundo sentido.

Mi premisa central es que clasificar relaciones E-S observadas en organismos biológicos constituye un aspecto central del problema mente-cuerpo y de la filosofía de la mente en general, independientemente de la posición específica que adoptemos. En efecto, clasificar constituye una actividad por demás ubicua y natural. Entre muchas otras cosas, y al igual que muchas otras especies, los humanos somos criaturas clasificadoras. Prácticamente cualquier problema científico o filosófico (incluyendo el problema mente-cuerpo), así como también cualquier solución propuesta, se origina, de una u otra forma, en nuestra habilidad para clasificar. Esta habilidad incluye la capacidad de discriminar sensorialmente entre, así como de actuar o reaccionar diferencialmente a, distintos

tipos de eventos ambientales y sus relaciones. Las clasificaciones, como productos lingüísticos de nuestra actividad de clasificar, representan herramientas organizadoras básicas en cualquier empresa intelectual. Empezamos a hacer filosofía o ciencia de algo desde el momento en que intentamos sistematizar nuestra búsqueda de explicaciones organizando, de acuerdo con ciertas categorías, las distintas instancias o casos de ese algo.

En el caso de la filosofía de la mente y el problema mente-cuerpo, las relaciones entre la conducta de los organismos y su medio ambiente, relaciones que pueden ser vistas como un tipo de relación E-S, constituyen los datos o materia prima fenoménica a partir de la cual experimentamos esa curiosidad, ese asombro que nos lleva a preguntarnos por qué los organismos (sean humanos o no) se comportan de la manera en que lo hacen. Ciertamente, cada día enfrentamos situaciones sociales en las cuales observamos y reaccionamos ante las conductas de otros y los contextos en los cuales éstas ocurren. Una persona riéndose en un funeral, llorando en una comedia o diciendo escuchar voces que nadie mas escucha son fenómenos que nos causan curiosidad y nos llevan a buscar explicaciones, quizás más que una persona llorando en un funeral, riéndose en una comedia, o diciendo escuchar voces que los demás escuchamos, aunque éstos últimos fenómenos también pueden constituir materia de investigación científica y reflexión filosófica, a pesar de que nos resulten menos sorpresivos. Nótese que los fenómenos de interés no son acciones, conductas o respuestas aisladas de sus respectivos contextos ambientales. No es la conducta de reír o de llorar *per se*, sino, más bien, la conducta de reír *en un funeral* o de llorar *en una comedia* lo que nos causa sorpresa. Para poder experimentar tal sorpresa, pues, debemos ser capaces de reconocer y reaccionar diferencialmente ante distintas *relaciones* entre conductas y ambientes, lo cual incluye discriminar tanto entre distintas formas de conducta como entre distintos tipos de ambientes.

Nos volvemos filósofos o científicos de la mente desde el momento en que nuestra curiosidad desemboca en una investigación sistemática que trasciende las situaciones cotidianas particulares que inicialmente la motivaron. Las clasificaciones resultantes de este proceso pueden ser muy diferentes de aquellas más intuitivas que sirvieron de punto de partida (tal y como la tabla periódica de los elementos químicos es diferente de la clasificación presistemática de los elementos en tierra, aire, viento y fuego). Pero nuestra habilidad para discriminar entre distintos tipos de comportamientos y ambientes es necesaria tanto para nuestros intentos iniciales de organizar esa materia prima fenómenica constituida por las relaciones ambiente-conducta que experimentamos día a día, como para aquellos fenómenos menos intuitivos, más sistemáticos, que construimos a lo largo de nuestra investigación filosófica y científica.

Todo lo anterior lleva a la idea de que, a la hora de hacer filosofía o ciencia de la mente, e independientemente de que seamos funcionalistas, materialistas, conexionistas, o dualistas, no podemos prescindir de los datos que nos proveen las relaciones entre la conducta de los organismos y su medio ambiente. Por supuesto, el énfasis sobre esas relaciones cambia de una a otra posición. Así, un materialista consecuente las verá como necesarias más no suficientes para una definición cabal del objeto de estudio de una disciplina de la relación mente- cuerpo, considerando la realización física particular de tales relaciones como un componente igualmente central de dicho objeto. Por su parte, un funcionalista a ultranza verá tales relaciones como el objeto de estudio propio de dicha disciplina, independientemente de sus realizaciones físicas particulares. Pero ninguno de ellos diría (creo yo) que las relaciones ambiente-conducta son irrelevantes y que podemos prescindir de ellas a la hora de hacer ciencia o filosofía de la mente. Ningún intento de formular o resolver el problema mente-cuerpo puede excluir relaciones ambiente-conducta. Metafóricamente, si vemos al problema como un rompecabezas, entonces clasificar relaciones ambiente-

conducta puede ser visto como una pieza de ese rompecabezas. Cualquier posición que adoptemos acerca de qué debemos investigar y cómo debemos hacerlo, a la hora de hacer filosofía o ciencia de la mente, debe tomar en cuenta dichas relaciones como parte integral de la investigación. Nuestra habilidad para clasificar relaciones ambiente-conducta, pues, se encuentra a la base de nuestras reflexiones filosóficas y nuestras descripciones y explicaciones científicas de cómo lo mental se relaciona con lo corporal. Sin esta habilidad, no habría filosofía ni ciencia de la relación mente-cuerpo. Por supuesto, el problema mente cuerpo es algo más que el problema de clasificar relaciones ambiente-conducta. Pero ser capaces de resolver este último problema es al menos necesario para cualquier intento de formular y de resolver el primero. Mi énfasis sobre las relaciones ambiente-conducta, pues, consiste en considerarlas como un componente fenoménico crítico para la filosofía y la ciencia de la mente, y la habilidad de clasificarlas como un aspecto central del problema mente-cuerpo, de nuevo, independientemente de que seamos dualistas, materialistas, funcionalistas o conexionistas. Sobre la base de este énfasis, mi uso del elemento MP responde a razones de claridad formal y simplicidad, con lo cual debe verse como un recurso metodológico, no como una defensa de filosofías particulares de la mente. El elemento MP resulta idóneo para la presente investigación debido a que, como mostraré más adelante, es un sistema clasificador típico.

Puesto sucintamente, el problema mente-cuerpo es el problema de determinar exactamente cómo ciertos eventos ambientales, ciertos eventos conductuales, y ciertos procesos mediadores y sus realizaciones físicas se relacionan entre sí. Un primer paso hacia una solución del problema reside en obtener clasificaciones teóricamente efectivas de las relaciones ambiente-conducta o, más genéricamente, relaciones E-S. El escenario virtual que propongo pretende simular (haciendo, por supuesto, todas las salvedades del caso) a un organismo que juega el papel de sujeto observador o cognoscente (e.g., un humano) observando a otro organismo del mismo

tipo o especie en el papel del objeto observado (e.g., otro humano). En el escenario particular, ambos, sujeto y objeto, son elementos del tipo MP. Mi objetivo es determinar las condiciones que debe satisfacer el sujeto para clasificar las relaciones E-S del objeto. Para ello, se hace necesario caracterizar detalladamente la estructura de un elemento MP.

El elemento MP es un sistema que posee un número de sensores conectados a una unidad de procesamiento. Los sensores reciben señales de entrada que representan estímulos del medio ambiente circundante y las transmite a la unidad de procesamiento, la cual ejecuta ciertas operaciones sobre dichas señales y retorna una señal de salida que representa la respuesta del elemento a las mismas. El efecto de una señal de entrada dependerá no sólo de su magnitud, sino también de la fuerza con la cual el sensor activado esté conectado a la unidad de procesamiento. En la jerga conexionista, dicha fuerza es cuantitativamente representada a través de un *peso*, mientras que la respuesta del elemento representa su *activación*. El funcionamiento del elemento se describe matemáticamente en términos de una regla que determina la respuesta del elemento en un momento particular (aquí, el tiempo es conceptuado como una variable discreta), como una función de las señales de entrada y de los pesos. Esta regla es una función condicional binaria que especifica dos estados posibles, a saber, '1' (o 'activado' o 'encendido') y '0' (o 'desactivado' o 'apagado'). El elemento puede encontrarse en uno y sólo uno de esos dos estados en cualquier momento en el tiempo. El elemento se activará si la combinación lineal de las señales de entrada y sus pesos correspondientes sobrepasa una magnitud conocida como *umbral*. De lo contrario, el elemento permanecerá desactivado.

Las propiedades del MP sobre las cuales deseo concentrarme tienen que hacer, por una parte, con su capacidad sensorial, definida como el número $n$ de sensores que posee, cada uno de los cuales puede estar activado (estado '1') o desactivado (estado '0'), y por otra, con su capacidad conductual o de respuesta, definida como el conjunto $R = \{1,0\}$ de

estados posibles. Sobre esta base, podemos definir el universo E-S de un elemento MP como el conjunto de todas las relaciones E-S que le son *lógicamente posibles*, aún cuando no pueda implementarlas funcionalmente. Dado el carácter binario de las activaciones, tanto del elemento mismo como de sus sensores, el tamaño del universo E-S viene dado por $2^{2n}$, el cual crece exponencialmente a medida que la capacidad sensorial del elemento crece aritméticamente.

Para mis propósitos en el presente trabajo, la utilidad del elemento MP no sólo reside en su simplicidad, sino también en su habilidad para discriminar. Decimos que un sistema discrimina si responde diferencialmente a distintos patrones de entrada, donde un patrón de entrada se define como un vector de estados de activación de los sensores del elemento en un momento determinado. Puesto que el elemento MP responde de manera dicotómica, el único tipo de discriminación que puede lograr adquiere la forma de una partición entre instancias de dos tipos diferentes de patrones de entrada. Si los patrones por clasificar son linealmente separables, entonces el elemento podrá clasificarlos. Esta capacidad clasificatoria del elemento MP ha sido útil en ciertas aplicaciones de ingeniería que requieren la automatización de ese tipo de ejecución (e.g., reconocimiento de patrones). ¿Pero qué podemos decir acerca de un elemento MP que confronta el problema de clasificar las relaciones E-S de otro elemento MP? ¿Qué condiciones debe satisfacer un elemento MP para ser capaz de clasificar las relaciones E-S de otro elemento MP? Con estas preguntas llegamos al núcleo de este trabajo.

Considérese, de nuevo, el escenario virtual descrito al principio. Sean $MP_K$ y $MP_O$ dos elementos del tipo MP tales que $MP_K$ juega el papel de sujeto observador de $MP_O$ y $MP_O$ juega el papel de sujeto observado por $MP_K$. Como punto de partida, supóngase que $MP_K$ enfrenta el problema de clasificar el universo E-E de $MP_O$ completamente y con un grano máximamente fino, de tal manera que cada instancia de ese universo define un taxón. ¿Puede $MP_K$ resolver este problema? ¿Bajo qué condiciones?

Para empezar por el lado sensorial de la cuestión, es claro que $MP_K$ debe por lo menos ser capaz de sensar las mismas señales de entrada que son sensadas por $MP_O$. Es decir, $MP_K$ debe por lo menos poseer la misma capacidad sensorial de $MP_O$. Por consiguiente, $MP_K$ debe por lo menos ser capaz de implementar to funcionalmente los mismos tipos de relaciones E-S que $MP_O$. Pero además, $MP_K$ debe también ser capaz de observar la conducta o de sensar las señales de salida de $MP_O$. La señal de salida de $MP_O$, pues, debe funcionar como una señal de entrada para $MP_K$. Si $MP_K$ no pudiera detectar la señal de salida de $MP_O$, entonces el primero no sería capaz de resolver el problema de clasificar las relaciones E-S del segundo. Por lo tanto, para poder realizar la tarea en cuestión, la capacidad sensorial de $MP_K$ debe ser mayor (en por lo menos un sensor) que la de $MP_O$. En general, si $n$ denota el número de sensores de $MP_O$, entonces $MP_K$ debe poseer al menos $n + 1$ sensores. No hay manera de que $MP_K$ enfrentar la tarea sin ese sensor adicional. Y esta diferencia mínima produce un incremento sustancial en el universo E-S de $MP_K$.

Tal incremento se hace más pronunciado si tomamos en cuenta el aspecto conductual, de salida o de respuesta. Conductualmente, el elemento MP es capaz de dar sólo dos tipos de respuesta. Sin embargo, el universo E-S de un elemento MP de, por ejemplo, dos sensores tendría un total de 16 relaciones distintas que serían lógicamente posibles. Una clasificación de este universo, entonces, consistiría de 16 categorías o taxones. Para que $MP_K$ pueda lograr tal clasificación, tendría que tener una capacidad conductual de tantos estados diferentes de activación como relaciones constituyan el universo E-S de $MP_O$, lo cual hace aún más numeroso el universo E-S de $MP_K$. Pero aunado a este incremento, un cambio conceptualmente más profundo ha ocurrido. Puesto que la capacidad conductual requerida rebasa la del elemento MP, $MP_K$ tendría que dejar de ser un elemento del tipo MP para pasar a ser un elemento o sistema conexionista de un tipo o especie diferente, un sistema cuya regla de activación permita por lo menos 16 estados de activación posibles.

Si juzgamos la *complejidad* de un sistema en términos del tamaño de su universo E-S, entonces podemos concluir que para que $MP_K$ sea capaz de clasificar el universo E-S de $MP_O$, $MP_K$ tiene que ser funcional y, por tanto, estructuralmente más complejo que $MP_O$. Si generalizamos esta conclusión a cualquier tipo de sistema obtenemos la siguiente tesis:

**Un sistema *K* sólo puede clasificar el universo E-S de otro sistema *O* de menor complejidad.**
o, en su forma negativa,
**Un sistema *K* no puede clasificar el universo E-S de otro sistema *O* de igual o mayor complejidad.**

Esta generalización, por supuesto, requiere de un análisis filosófico que no puedo llevar a cabo aquí. Si esta tesis soporta tal análisis, entonces las implicaciones para nuestras posibilidades de solucionar el problema mente–cuerpo pueden ser profundas. En efecto, si la solución del problema, aplicado a los humanos, depende de nuestra capacidad de entendimiento, y si tal capacidad (sea lo que sea) depende de nuestra capacidad de clasificar el universo E-S humano, y si los humanos somos más o menos igualmente complejos unos con respecto a otros, entonces un humano no puede clasificar el universo E-S de (y, por tanto, entender) otro humano. La implicación principal para una disciplina como la psicología es que nunca podremos alcanzar un entendimiento cabal de la conducta humana. El único tipo de psicología que seremos capaces de construir es una psicología de sistemas considerablemente más simples que nosotros.

Instituto de Psicología
Universidad Central de Venezuela

## REFERENCIAS

McCulloch, W. S., & Pitts, W. H.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 1943, pp. 115-133. Reprinted in W. S. McCulloch, *Embodiments of mind* , Cambridge, MA: The MIT Press, 1988, pp. 19-39.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 63–87

ANTONI GOMILA BENEJAM

# EXPERIMENTOS MENTALES EN CIENCIA Y EN FILOSOFIA

**RESUMEN:** Una característica destacada, si bien no muy atendida, del pensamiento científico es el recurso a los experimentos mentales. En este trabajo propongo una concepción de los procesos mentales que involucran, a partir del análisis de algunos de los experimentos mentales más notables en la historia de la ciencia. Frente al empirismo dominante en adquisición del conocimiento y cambio conceptual, que no deja espacio alguno para su relevancia epistémica, sostengo que los experimentos mentales pueden constituir una forma genuina de razonamiento dirigido al mantenimiento de la consistencia, que puede dar lugar al cambio conceptual. Para ello, se desarrolla un modelo general de la naturaleza del cambio conceptual, que permite entender cómo puede resultar de un proceso reflexivo. Frente al externismo semántico extremo, se propone que la competencia conceptual implica disponer de alguna concepción mínima implícita de la clase de referencia, que puede explicitarse mediante el planteamiento de situaciones contrafácticas como las planteadas en los experimentos mentales, y de este modo, pueden ponerse de manifiesto inconsistencias en tales concepciones y, por consiguiente, vías de reforma, que pueden generar cambios en la propia delimitación de la clase de referencia. Finalmente, se propone que la misma clase de competencia y de proceso están implicados en los experimentos mentales en filosofía, en lo que constituye una concepción semántica del conocimiento a priori.

**Palabras Claves:** Semántica, referencia y cambio conceptual.

**ABSTRACT:** A remarkable feature of scientific thinking, though one usually unattended to, is thought experimentation. In this paper, I propose to take a look at real thought experiments, in order to reach a characterization of the sort of thinking processes they involve. Against the dominant empiricist views on knowledge formation and conceptual change, that downplay or

explain away the import of thought experiments in scientific thinking, it is argued that thought experiments constitute a genuine form of reasoning aimed at coherence keeping, which can be instrumental to conceptual change. To this aim, a model of conceptual competence is presented, understood as a modification of standard semantic externalism, which distinguishes among the concept, the implicit conception and its explicit formulation. Thought experiments, in this view, are a means to making explicit the conception, which can turn out to involve inconsistencies. Modifying this understanding may involve a change in the class of reference of the concept. Finally, it is claimed that much the same process is involved in thought experiments in philosophy, so that no special faculty of conceptual intuition needs to be put forward to account for a priori knowledge, which is accordingly understood as grounded in our conceptual competence.

**Keywords:** Semantics, reference, conceptual change.

## *1. Introducción*

Una de las discusiones más apasionadas que ocupó a la física de los siglos XIV y XV fue la de la posibilidad del vacío. El vacío, según la física aristotélica, es imposible, pero los críticos nominalistas de Aristóteles sostenían que sí era posible crearlo. La base de su afirmación era el siguiente experimento mental. Supongamos que se llenara completamente una jarra con agua, se sellara y se expusiera al frío invierno; el agua, al helarse, dejaría un espacio vacío en la parte superior. Los aristotélicos replicaron que, en tal caso, o bien el hielo desprendería vapores que llenarían ese espacio, o bien simplemente la jarra se rompería al contraerse el hielo. La respuesta de los defensores del vacío consistió en proponer la sustitución de la jarra por una esfera de hierro, cuya dureza impediría la implosión. En fin, el debate se fue complicando con mayores sutilezas y refutaciones. Sin embargo, ambas partes pasaron por alto un detalle fundamental: que el agua al helarse se expande, no se contrae. Además, la demostración de la existencia del vacío se consiguió años más tarde con el termómetro de mercurio y la construcción de la bomba de aire –no mediante discusiones de experimentos mentales.

La moraleja que suele extraerse de este tipo de historias consiste en la reafirmación de que es la experimentación real, efectiva, la que guía el progreso científico, mientras que la especulación apriorística, conceptual, no es más que un juego autosostenido que no sólo no contribuye al avance del conocimiento, sino que lo retarda. Sólo de la experiencia metódica puede obtenerse conocimiento, no de la reflexión intelectual. Lo cual, por cierto, deja en un lugar epistemológico especial, por no decir en mal lugar, a la propia filosofía, cuya única experimentación, en principio, puede ser mental.

Sin embargo, no todos los ejemplos de experimentos mentales que encontramos en la historia de la ciencia son tan desalentadores. Al contrario, justamente algunos de los grandes nombres de la ciencia –Galileo especialmente, pero también Newton, Darwin y Einstein– están asociados con experimentos mentales decisivos para el cambio teórico. Su utilización de experimentos mentales no consiste en un mero recurso expositivo de sus nuevas teorías, sino que constituyen un medio decisivo para la desacreditación de las anteriores y la motivación de las suyas. Todo lo cual nos lleva a plantear una serie de preguntas:

- ¿en qué consiste un experimento mental? ¿son todos del mismo tipo?
- ¿cómo funcionan? ¿aportan realmente conocimiento? ¿en virtud de qué?
- ¿son sólo ilustraciones de ideas obtenidas por medios propiamente experimentales?
- ¿cuándo cabe recurrir a ellos? ¿cuando la experimentación efectiva no está al alcance o cuando lo que se dirimen son cuestiones conceptuales?
- ¿cuál es su relación con los experimentos mentales que encontramos en filosofía?

No voy a responder a todas ellas en detalle ni directamente. Quisiera presentar solamente una forma de entender los experimentos mentales, al menos aquellos experimentos mentales que más llaman la atención en la historia de la ciencia por su influencia en el cambio de teoría. Mi propuesta es concebirlos como medios para la explicitación de nuestra

concepción de los conceptos teóricos; en este sentido, se basarían principalmente en nuestra competencia semántica/conceptual. Esa explicitación puede poner de manifiesto consecuencias inconsistentes, o posibilidades que los conceptos disponibles no alcanzan, y aun el medio de cambiar nuestros conceptos para evitarlas. Por supuesto, el alcance de esta explicación pretende ser general; no hay métodos que garanticen el éxito en la ciencia, por lo que los experimentos mentales exitosos no lo son por basarse en un mecanismo especial.

En realidad, debo confesar que mi interés por los experimentos mentales está subordinado a mi interés por la teoría de los conceptos. Dar cuenta de la posibilidad de que los experimentos mentales contribuyan al conocimiento, vía la reforma o el cambio conceptual, presupone disponer de una teoría de los conceptos en la que sea posible el cambio conceptual como resultado de la reflexión sobre posibilidades contrafácticas. Y eso es algo que no parece encajar fácilmente en el programa anti–individualista o externista dominante actualmente, en el que el interés se centra en la individuación conceptual, más que en la competencia. Dicho de otro modo, la existencia de experimentos mentales fructíferos, en mi opinión, sirve para poner de manifiesto los límites del enfoque externista dominante en la teoría de los conceptos. Tener en cuenta los experimentos mentales obliga a suplementar las condiciones de posesión de conceptos con condiciones de competencia conceptual, más allá de las habituales condiciones lingüísticas o sociales. O mejor dicho, a enriquecer las condiciones de posesión para que constituyan al mismo tiempo condiciones de competencia. En este sentido, voy a introducir una distinción entre conceptos, concepciones y su explicitación consciente, que luego aplicaré al caso de los experimentos mentales.[1]

---

[1] En otro trabajo (en prensa, a), desarrollo el planteamiento para el cambio conceptual en general. Los experimentos mentales constituyen solamente una vía, ocasional, para el cambio conceptual.

De este planteamiento se desprende, además, creo, una concepción de los experimentos mentales en filosofía como análogos a los de la ciencia. En parte, por el impacto mismo del anti–individualismo, con su insistencia en que para tener un concepto no hace falta conocer "la esencia" de aquello a lo que se refiere. Frente a las pretensiones de alcanzar un conocimiento necesario a priori del análisis conceptual ortodoxo, herederas del racionalismo moderno y del idealismo absoluto que se expresa en el lema "lo real es racional, lo racional es real", el externismo semántico implica un visión más moderada de lo que se puede alcanzar mediante los experimentos mentales y la reflexión en general. La cuestión no es negar la posibilidad de conocimiento a priori, sino entenderlo, no como la ejercitación de una intuición especial, sino como la reflexión sobre la propia competencia conceptual.

## *2. Experimentos mentales en la ciencia*

Para situar la discusión, lo mejor será empezar con algunos ejemplos de experimentos mentales. En el ámbito de la ciencia, voy a presentar dos de los más citados e impactantes, uno de Galileo y otro de Einstein.

Según la teoría aristotélica, cuanto más pesado es un cuerpo mayor será su velocidad natural de caída. Esto parece intuitivamente obvio: si dejamos caer esta hoja de papel y el bolígrafo con el que estoy escribiendo, el bolígrafo llegará antes al suelo. Por otra parte, Aristóteles también notó que cuando a un objeto más rápido se le une otro más lento, el más rápido pierde velocidad. Así, por ejemplo, si a una liebre le atamos una tortuga, la liebre no podrá ir tan deprisa.

El experimento mental de Galileo consiste en combinar estos dos principios de la física aristotélica para concluir que llevan a una contradicción. Supongamos que a una gran piedra le conectamos mediante un medio sutil otra piedra más pequeña durante su caída. Por el primer principio cabe esperar que el compuesto de las dos piedras caiga a mayor velocidad que cualquiera de las dos piedras por separado, ya que el compuesto es más pesado. Por el segundo principio, sin em-

bargo, resulta que su velocidad debe ser menor que la de la piedra grande sola, ya que la pequeña la frena en su caída. Ahora bien, no puede ser que caiga más y menos rápido al mismo tiempo. Conclusión: hay algo mal en la teoría aristotélica. La solución, para Galileo, está en que ambas piedras caen a la misma velocidad. Para que ello tenga sentido es preciso separar peso de velocidad de caída, y al mismo tiempo, atribuir a factores de rozamiento y resistencia del medio en que tiene lugar la caída la obviedad de que de hecho no caen al mismo tiempo. En estos dos aspectos, sobre todo el primero, consiste el esfuerzo teórico de la física galileana.

Los escritos de Galileo, en especial los Diálogos, están repletos de experimentos mentales, algunos de los cuales tienen la misma elegancia y fuerza dialéctica que éste. Quizá sólo Einstein está a la altura de Galileo en este respecto, utilizando profusamente los experimentos mentales como recurso metodológico a lo largo de su carrera. Quizá el más conocido sea el del ascensor cósmico. Hay que situarlo en el contexto de las tensiones entre la mecánica clásica y la electrodinámica, que su teoría especial de la relatividad había contribuido a resolver, pero sólo con respecto al movimiento inercial. Esa solución consiste en relativizar la noción de simultaneidad a un marco de referencia espacio–temporal. Hacía falta hacer algo parecido con respecto al movimiento acelerado. La vía de solución del conflicto, la equivalencia de la masa inercial y masa gravitacional, se encuentra en la formulación de este experimento mental.

Einstein plantea la posibilidad de un ascensor que está siendo elevado por una fuerza constante en el espacio. Dentro del marco de la mecánica clásica hay que presuponer que el ascensor está en un sistema inercial, un sistema con respecto al cual puede hablarse de movimiento absoluto, y por tanto, su movimiento puede ser considerado como uniformemente acelerado. Sin embargo, para un observador que esté dentro del ascensor no está claro que las cosas tengan que ser así. Podría creer igualmente que lo que ocurre en realidad es que el ascensor está estacionario, sometido al campo gravitatorio

de una masa próxima. Dicho de otro modo, no es posible distinguir, en tal situación, entre una posibilidad y otra. Para ilustrarlo, Einstein plantea que un rayo de luz entre por un orificio lateral del ascensor; o más intuitivamente, que se deje libre un objeto en el interior. Es de suponer que el objeto caería, pero el observador interior no podría decidir si se trataba de un fenómeno gravitacional o resultado del movimiento acelerado del ascensor.

No se trata en este caso de una contradicción directa, sino de detectar un "punto ciego" en la teoría clásica. Para Einstein, el propio experimento mental indica cuál es la vía para enmendarla: prescindir de las nociones clásicas de sistema inercial y movimiento absoluto, y partir de la equivalencia de masa inercial y masa gravitacional, es decir, asumir que las dos posibilidades en realidad son sólo una. Este es el punto de partida de la teoría general de la relatividad.

No todos los experimentos mentales que encontramos en la ciencia han sido tan decisivos como éstos que hemos presentado, desde luego. Ni tienen la misma finalidad. A veces se presentan como experimentos mentales situaciones que plantean simulaciones mentales de "casos límite" o de órdenes de magnitud diferentes al conocido. Pero los que son como éstos contribuyen sin duda al progreso científico, al avance teórico. A primera vista, presentan ciertas características distintivas: integran una dimensión destructiva o crítica respecto a una teoría establecida, y una dimensión constructiva, al sugerir la dirección para la nueva teoría, aunque esta dimensión está vinculada al diagnóstico del problema detectado. No consisten en aportar nueva información -¿cómo podrían conseguirlo?-, pero son informativos, al poner de manifiesto que ciertas posibilidades inaceptables, o ignoradas, se siguen de esa teoría establecida, y que para tenerlas en cuenta, o remediarlas, es preciso modificar los conceptos manejados. No necesitan ser reproducidos efectivamente –muchas veces es completamente imposible hacerlo– para ser valiosos. No se limitan, además, a lo meramente posible desde un punto de

vista nomológico, como a veces se ha sugerido[2], ni se basan en la experiencia común a la comunidad científica[3]: el ascensor de Einstein es claramente una imposibilidad física y no se basa en la experiencia común. Es más, el experimento mental contribuye a establecer tales posibilidades nomológicas, fijadas por la nueva teoría.

Ahora bien, las conclusiones obtenidas no están fuera de toda duda, al contrario, son falibles. No proporcionan una vía de acceso a un conocimiento privilegiado, sino que su aporte depende directamente de las críticas hechas a la teoría anterior. Es preciso fijarse en los experimentos mentales fructíferos para reconocer su papel en la ciencia, pero no hay un algoritmo para generarlos, ni un método para garantizar su éxito. El experimento mental no basta, por sí mismo, para dar lugar a un cambio teórico. Pero todo esto no es algo especial, sino que es propio de la dinámica científica: hasta que no hay una alternativa desarrollada no se abandona la teoría establecida, a pesar de las posibles anomalías –en el sentido de Kuhn– que la afecten. Lo importante es que los experimentos mentales puedan contribuir al cambio teórico, y que de hecho han contribuido en los cambios teóricos más importantes. El reto que plantean es explicar como es ello posible.

### *3. Enfoques disponibles acerca de los experimentos mentales*

Si el análisis anterior está bien encaminado, entonces ninguna de las diversas formas de considerar los experimentos mentales, más o menos inspiradas en la confrontación

---

2 Wilkes, K., *Real people. Personal identity without thought experiments,* Oxford, Clarendon Press, 1988, sostiene que los experimentos mentales son aceptables en la ciencia porque consideran posibilidades nomológicas, pero inaceptables en filosofía porque implican posibilidades metafísicas con respecto a las cuales carecemos de criterios de plausibilidad.

3 Kuhn, T., "A function for thought experiments in Einstein's work" en *The Essential Tension,* Chicago, The University of Chicago Press, 1977, ha sostenido que la eficacia de los experimentos mentales se basa en partir de la experiencia común a la comunidad científica, además de en poner de manifiesto la tensión conceptual. Pero, como el ejemplo de Einstein ilustra, ello no tiene porqué ser así. Kuhn asume un constructivismo psicológico con respecto a los conceptos que es en sí mismo problemático.

racionalismo–empirismo tradicional, puede resultar plenamente satisfactoria. Dentro de la tradición empirista dominante en la filosofía de la ciencia se ha tomado al pie de la letra la idea de que todo conocimiento tiene su origen en la experiencia (como si no hubiera formas mediatas de experiencia), y por tanto, que la mera reflexión no constituye un medio apropiado para proporcionar conocimiento. Se ha sostenido, por ello, que los experimentos mentales carecen de interés genuino, que son más bien una forma camuflada de presentar un argumento [4], o bien, de tener interés, corresponderían al contexto de descubrimiento, no de la justificación; por tanto, los experimentos mentales deben situarse al margen de la lógica de la ciencia [5].

Por su parte, los defensores del racionalismo encuentran en los experimentos mentales su mejor apoyo para sostener que la mera reflexión es capaz de alcanzar verdades científicas. Los estudios de Koyré sobre Galileo, por ejemplo, insisten en mostrar la irrelevancia de la experimentación efectiva para la teoría galileana, cuyo sostén se encuentra más bien en toda una serie de experimentos mentales [6]. Los escritos del Einstein maduro también dan la impresión de que no le importaban demasiado las eventuales confirmaciones empíricas de sus teorías. Recientemente, el racionalismo epistemológico ha sido defendido por Brown [7] por analogía con el raciona-

---

4 Norton, J., "Thought experiments in Einstein's work, e, *Thought Experiments in Science and Philosophy*, T. Horowitz and G. Massey (Eds.), Savage, MD Rowman and Littlefield, 1991

5 Hempel, C., *Aspects of Scientific Explanation,* New York, Free Press, 1965. Dentro del empirismo decimonónico, la anomalía es Mach, quien se interesó en primer lugar por el papel de estos experimentos en la ciencia, y acuñó el término "Gedankenexperiment" para referirse a ellos (Mach, 1905). Su interesante propuesta remite a la evolución como un medio de transmisión de conocimiento, originado en la experiencia de generaciones anteriores. La epistemología evolutiva de Mach es insatisfactoria por su inespecificidad, pero encierra interesantes intuiciones que han influido en la propuesta defendida aquí. Un problema añadido es su dependencia de la psicología introspeccionista y subjetivista.

6 Koyré, A., "Galileo's Treatise 'De motu gravium':: the use and abuse of imaginary experiment", en *Metaphysics and Measurement*, Londres, Chapman and Hall, 1968.

7 Brown, J. R., *The laboratiry of the mind. Thougth experiments in the natural*

lismo matemático o platonismo: las leyes de la naturaleza, como las verdades matemáticas, habitarían un mundo ideal que, en ocasiones, somos capaces de captar mediante la intuición. Este neo–racionalismo, sin embargo, pasa por alto varias cosas. En primer lugar, que la experimentación efectiva es insoslayable en la actividad científica. El papel de los experimentos mentales no puede ser el de sustituirla. Además, la noción de conocimiento a priori que conlleva este planteamiento es completamente problemática; mientras que con respecto al conocimiento matemático, por su abstracción, puede tener sentido, no hay modo de asegurar a priori el acceso a verdades empíricas, ni siquiera a los conceptos teóricos apropiados. Lo que es peor: los experimentos mentales son falibles, como hemos visto, lo cual encaja mal con esa supuesta intuición racionalista.

Kuhn propuso hace unos años, dentro de su concepción sociológica de la ciencia, que la función de los experimentos mentales era propiciar una reorganización de la información que llevara a la formulación de nuevos conceptos, en base a la confrontación con la experiencia de la comunidad científica [8]. Aunque tiene el mérito de haber destacado que la detección de posibilidades anómalas era la principal contribución de los experimentos mentales –si bien Kuhn habla de conceptos contradictorios, no de posibilidades propiamente–, su insistencia en la inconmensurabilidad de teorías, hace imposible entender cómo esta misma dimensión crítica puede conllevar el germen de la nueva propuesta. Es más, un experimento mental presupone en el fondo la posibilidad de comparar una teoría con otra. Sin embargo, al margen del marco epistemológico kuhniano, la presente propuesta puede considerarse en la línea abierta por su trabajo.

Recientemente, los enfoques socio–cognitivos de la ciencia, surgidos en parte para "solucionar" los problemas que el sociologismo estricto no puede ni siquiera plantear, han sugerido que los experimentos mentales funcionan porque son

*sciences,* Londres, Routledge, 1991.

8 Kuhn, *op. cit.*

como experimentos efectivos hechos en la cabeza. Del mismo modo que uno puede "contar" en la imaginación el número de ventanas de su casa, reproduciendo mentalmente el procedimiento que seguiría si fuera a contarlas efectivamente, uno puede "experimentar" con la imaginación, ya que nuestros "modelos mentales" incorporan no sólo conceptos sino también procedimientos. Los experimentos mentales son simulaciones mentales.[9] Esto puede ser cierto e interesante, pero a mi modo de ver, tiene que ver con la crítica a un modelo exclusivamente proposicional del conocimiento científico, que sólo tiene en cuenta el "saber qué" pero no el "saber cómo", y en este sentido, es sólo otra forma de decir que los experimentos mentales no aportan información nueva. Pero no da cuenta de lo realmente fascinante de los experimentos mentales que nos interesan: que abren una vía para una nueva forma de pensar, para una nueva teoría, para nuevos conceptos. En general, las simulaciones no son capaces de hacer esto.

Por otra parte, mientras que la oposición tradicional empirismo–racionalismo discrepa sobre las virtudes epistémicas de la reflexión, ambas posiciones coinciden en que hay una continuidad entre experimentos mentales en ciencia y filosofía; mientras que el empirismo los rechaza en ambos, el racionalismo los acepta en ambos. Pero para este enfoque sociocognitivo, parece plantearse un abismo entre los experimentos mentales de un tipo y de otro, lo cual tampoco es satisfactorio, como trataré de mostrar en la sección final.

Frente a estas propuestas, quisiera elaborar la idea de que los experimentos mentales son posibles porque se basan en la competencia semántica/conceptual. Por supuesto, intervienen también otras capacidades cognitivas, como las inferenciales y las intuiciones modales de posibilidad, imposibilidad y necesidad. Diré también algo con respecto a estas otras, pero

---

9 Gooding, D., "The Procedural Turn" en Giere, R. (Ed.) *Cognitive models of science Minnesota Studies in the Phylosophy of Science*., Minneapolis, Minnesota U.P., 1992. Nersessian, N., "How do scientists think? Capturing the dynamics of conceptual change in science", en Giere, *op, cit.*

me parece que la perspectiva más relevante es la de ver los experimentos mentales como medios para explicitar la concepción implícita de los conceptos teóricos y sus consecuencias; es decir, están basados en nuestro conocimiento semántico/conceptual, en nuestra competencia, en lo que entendemos con tales conceptos.[10]

### *4. Una teoría de la competencia conceptual*

En los términos planteados, no obstante, es probable que la propuesta formulada genere cierto escepticismo, por su aroma verificacionista, o por su falta de delimitación de los niveles epistemológico y semántico. Lo que hace falta, por tanto, es formular con carácter previo la concepción de la competencia semántica/conceptual que sostiene tal propuesta, una concepción que se sitúa dentro del ámbito del anti–individualismo semántico contemporáneo, pero que parte del diagnóstico de que tal anti–individualismo cuenta con una noción empobrecida de competencia semántica/conceptual.

Lo que el anti–individualismo, en las diversas versiones de Kripke, Donnellan, Putnam o Burge, ha puesto de manifiesto es que un hablante puede usar con su significado un término sin tener un conocimiento correcto y completo de la naturaleza esencial de aquello a lo que se refiere ese término. La razón de ello, más allá de los diversos modos en que han tratado de desarrollarla, radica en que las condiciones para la posesión de un significado/concepto no son epistémicas, frente a la comprensión estándar de Frege; pueden ser históricas, causales, sociales, o una mezcla de ambas. El énfasis en estos aspectos externistas ha llevado a relegar a un segundo plano el papel que deba jugar el conocimiento, correcto o no,

---

10 Sorensen, R., *Thought experiments,* Oxford, Oxford U. P., 1992, ha propuesto la idea de que los experimentos mentales son refutadores aléticos, si bien pone más énfasis en el análisis lógico de los experimentos mentales que en los conceptuales y psicológicos. Pero puede verse como el complemento de lo que aquí se plantea, en relación a esas otras capacidades involucradas, en particular lo que podría llamarse el sistema de mantenimiento de la consistencia: como parte de los mecanismos dedicados a evitar contradicciones internas en nuestro conocimiento.

de la extensión del término en la teoría de los conceptos.[11]

Sin embargo, me parece que es claro que las condiciones de posesión de un concepto tal como se caracterizan por parte de los enfoques externistas dominantes no equivalen a una caracterización de la competencia semántica. Un modo de ponerlo de relieve consiste en pensar en lo que puede considerarse como una restricción relevante para la atribución conceptual: que la atribución canónica de conceptos a un sujeto (según nuestra mejor teoría anti–individualista) debe coincidir con los que el propio sujeto puede atribuirse a sí mismo, con respecto a los conceptos que es capaz de usar de modo consciente, a través de sus preferencias lingüísticas, por ejemplo. Permítaseme ilustrar lo que quiero decir con un par de ejemplos, quizá algo esquemáticos.

Supongamos que en uno de mis paseos por Buenos Aires me encuentro con una situación que me parece de emergencia por un accidente. Alguien, muy sofocado, se me acerca, gritando "pichincha, pichincha"; atribulado, me pongo a correr repitiendo ese grito, en la confianza de que quien lo oiga va a ser capaz de proporcionarla (o llamarle, o hacer algo; por mi parte, ignoro si es un nombre propio o común, si es denotativo o).

La clave del ejemplo radica en que se trate de una palabra propia del sociolecto bonaerense, pero ausente en el mío; en tales circunstancias, mi audiencia no tendría ninguna dificultad en entender el mensaje que yo me he limitado a repetir. Mi uso de la palabra sí es significativo y referencial –pues la referencia, como sostiene el externismo, depende de factores sociales y normativos que me superan–, y mantiene su

---

[11] Así, por ejemplo, aparece en la hipótesis de la división del trabajo lingüístico de Putnam, al remitir a los expertos como poseedores del conocimiento correcto y completo de la naturaleza esencial de la extensión, y por tanto, capaces de determinar si algo efectivamente está en la extensión, lo que, sorprendentemente, parece reintroducir la idea fregeana de que es el sentido, lo conocido, lo que determina la referencia, aunque sea solamente para una selecta minoría. (Es notable el hecho de que Putnam no se plantee siquiera si esta especialización lingüística debe corresponder a cada lenguaje.) Burge, en cambio, es quien más sensible ha sido a lo que se plantea aquí, como se verá en breve.

significado establecido. Pero parece inapropiado atribuirme competencia semántica en tal término. Yo, desde luego, rechazaría tal atribución: ni sé lo que significa ni a qué se refiere, ni en ese contexto concreto ni en general.

Otro caso. Supongamos que tiene lugar una radiación cósmica que pasa inadvertida a todo el mundo, que tiene por efecto la transformación del hidrógeno, cuando está combinado con el oxígeno formando agua, en helio. Dicho de otro modo, no hace falta desplazarse hasta la Tierra Gemela para pensar en una situación en la que el agua no es $H_2O$. Si las teorías externistas fueran correctas, habría que atribuirnos un cambio conceptual que, de nuevo, no aceptaríamos como una descripción correcta de nuestros pensamientos, al menos hasta tanto no advirtiéramos de algún modo el cambio.

A lo que trato de apuntar es al envés de la argumentación externista: podemos aceptar que no hace falta un conocimiento correcto y completo del referente para ser capaz de referirnos a él, pero quizá sea excesivo prescindir de todo conocimiento al respecto. Quizá hace falta un mínimo de competencia semántica para que pueda atribuirse a alguien un concepto. Tener un concepto, dado que los conceptos son los constituyentes de nuestros pensamientos, no puede ser un fenómeno totalmente externo –como estar a 42º de latitud sur o rodeado de descendientes de europeos–.

Eso no tiene porque suponer que lo que uno sabe determina qué concepto tiene uno, de modo que sea imposible distinguir entre una comprensión errónea de un concepto y una correcta de otro distinto. Al contrario, creo que una de las consecuencias valiosas del externismo semántico es señalar justamente la diferencia entre poseer un concepto y poseer una comprensión completa y correcta de la naturaleza de su extensión (su esencia, en la terminología aristotélica, aquello que garantizaba la *adequatio* entre conceptos y substancias). Esta primera diferencia es la que recojo con la distinción entre "conceptos" y "concepciones", lo que Burge ha denominado "el concepto" y "la explicación conceptual", o a nivel del lenguaje, "significado de traducción" y "significado

léxico".[12] A diferencia de Burge, sin embargo, por concepción entiendo no sólo la información de que se dispone sobre el concepto, esto es, aquello que lo pone en relación con los demás, y que permite explicar lo que uno entiende al respecto, sino también la competencia referencial, la capacidad de reconocer, de reidentificar, las diversas instancias del concepto como tales, al menos en algunas situaciones paradigmáticas. En tales situaciones, que podrán ser relativas a cada tipo de concepto, resulta criterial para reconocer competencia semántica a alguien que sea capaz de aplicar ese concepto en tal situación.

Además, Burge es ambiguo con respecto a la relación entre ambos.[13] En mi opinión, lo que señalan los ejemplos anteriores y la restricción apuntada es la necesidad de considerarlos como mínimamente relacionados, puesto que en el caso de experimentos mentales como los reseñados lo que ocurre justamente es una "reforma" extensional en virtud de las anomalías intensionales detectadas. Entender la naturaleza de esta relación va a resultar fundamental para entender cómo un experimento mental puede llevar a un cambio conceptual, esto es, a un cambio de concepto, y no sólo a un cambio de concepción (que carece del impacto, la repercusión y lo sorprendente de los grandes experimentos mentales en que nos hemos centrado). Creo que pueden apuntarse cuatro aspectos: a) sólo tenemos acceso a nuestros conceptos a través de las concepciones que tenemos de ellos, aunque sean incorrectas o incompletas; b) es preciso tener alguna concepción para tener el concepto –junto, por supuesto, la satisfacción de los factores externos señalados por el anti–individualismo–; c) no es preciso que haya alguien, un experto, perteneciente a la propia comunidad lingüística o no, que sí tenga lo que podríamos llamar una "concepción máxima", perfectamente adecuada y completa, para que el resto pueda poseer el con-

---

12 Ver, por ejemplo, Burge, T., "Concepts, definitions and meaning", *Metaphilosophy 24,* (1989), 1993.

13 Esto lo he tratado de argumentar en mi "Thought experiments and semantic competence", en prensa, b.

cepto; d) hay muchas formas de tener un concepto, es decir, diversas concepciones pueden ser del mismo concepto. La intuición de fondo que justifica este planteamiento parte de que sin una distinción entre verdades analíticas y sintéticas, no es posible restringir la competencia semántica al conocimiento de las verdades en virtud del significado, frente a las empíricas; esto es compatible, no obstante, con el hecho de que la gente trate ciertos aspectos de sus concepciones como definitorios del concepto.

Lo que resulta de ello es que para tener un concepto se requiere un mínimo de competencia conceptual, lo que refleja el carácter gradual de la competencia semántica/conceptual, y la ocasional indeterminación que afecta la tarea de adscribir conceptos, justamente en los casos en que la competencia es tan limitada, o se da una laguna importante, o domina el error, que resulta problemático justificar la posesión del concepto. Dicho de otro modo, se tiene el concepto en virtud de contar con cierta competencia mínima –cuánta, puede variar según el tipo de concepto, más o menos próximo al mesocosmos, más o menos ligado a un dominio especializado–, además de pertenecer a cierta comunidad lingüística y vivir en determinado mundo; lo cual no significa que el concepto se pueda analizar en base a la concepción, que el concepto se reduzca a la concepción, o que la concepción determine el concepto. El objetivo de la propuesta es enriquecer la intuición externista con la toma en consideración de los requisitos para la competencia conceptual, de modo que las condiciones para la posesión de un concepto sean también las condiciones para la competencia conceptual (mínima).

Establecida esta diferencia entre concepto y concepción, debemos introducir una tercera noción relacionada, la de "explicitación consciente" del concepto.[14] Una forma de en-

[14] Tomo esta tercera noción de Higginbotham, J. "How do I Know what my words mean?" en *Knowing one's own mind*, Oxford, C. Wrigth & B. Smith (Eds.), 1997, quien también añade las otras dos, pero desde supuestos semánticos muchos más clásicos, todavía ligados a la idea de verdades analíticas como las constitutivas de los análisis semánticos, y de la posibilidad de una

tender la noción de concepción que he presentado es como un "archivo", como el conjunto de información disponible, de capacidades perceptivas de reconocimiento asociadas, de procedimientos y saberes no proposicionales vinculados, al concepto[15]. En su mayor parte, por tanto, pueden ser implícitas. Como ya observó Agustín, y nos recordó Wittgenstein, todo el mundo sabe lo que es el tiempo, y sin embargo, resulta muy difícil de explicar. De hecho, es posible incluso que la explicitación consciente que alguien pueda hacer de un concepto no responda a su concepción implícita, creándose así un fenómeno de disonancia cognitiva. En general, no obstante, cabe esperar que la explicitación consciente refleje algunos aspectos al menos de la concepción, quizá los aspectos prototípicos o paradigmáticos. Por ello pueden producirse discrepancias y desacuerdos entre diferentes sujetos competentes, sin que sea plausible decir que, en realidad, una descripción correcta de la situación sería que no hay tal desacuerdo sino conceptos diferentes.[16] La captación de un concepto ni depende de conocer las condiciones necesarias y suficientes para que algo pertenezca a su extensión, ni se agota en su definición. Ilustraciones de este punto se pueden encontrar especialmente en el ámbito de la matemática: el mismo concepto de algoritmo de Turing, Church y Post, a pesar de sus diferentes caracterizaciones; el mismo concepto de infinitesimal de Leibniz y Newton, a pesar de sus definiciones incompletas, a falta de la noción de límite de una función en

---

"concepción adecuada" en el sentido de la pretensión clásica de captar en una definición la esencia de la clase de referencia. También Peacocke ha defendido la noción de "concepción implícita", Peacocke, C. "Implicit conceptions, understanding, and rationality", manuscrito, 1997, sugiriendo su diferencia con la explícita, también dentro de un enfoque que supone la posibilidad de analizar los significados, de definir los conceptos, por medio de condiciones necesarias y suficiente (las "condiciones de posesión" de Peacocke).

15 Este último aspecto es lo que Searle ha llamado el "trasfondo"; Searle, J., *Intentionality,* Cambridge, Cambridge 1983; Searle, J., *The rediscovery of the mind*, Cambridge, MS., MIT Press, 1992.

16 En realidad ésta es una consecuencia de la aplicación de la estrategia de Moore de la "pregunta abierta". Vd. Gibbard, A., *Wise choices, apt feelings,* Oxford, Clarendon Press, 1990. Introducción.

un punto[17]. Pero no es exclusivo de este ámbito: la noción de evolución por selección natural fue pensada independientemente por Darwin y Wallace; pero los ejemplos no se limitan a las convergencias exitosas, también a las discrepancias con sentido, donde el resultado histórico es la derrota de una explicitación consciente, incluso de las concepciones implícitas. En cualquier caso, la existencia de una concepción implícita es lo que explica la capacidad de reconocer como acertada una visión explícita proporcionada por otro, o de formular una uno mismo.

### *5. Los experimentos mentales como medios para la explicitación de la competencia conceptual*

A riesgo de ofrecer una imagen sesgada de los experimentos mentales, al relegar a un segundo plano otras capacidades cognitivas involucradas en los experimentos mentales, como el razonamiento hipotético o las intuiciones modales, quisiera sostener que los experimentos mentales consisten en herramientas dirigidas a explicitar las concepciones asociadas a un concepto, y que es por este motivo por lo que a veces pueden tener el éxito de los experimentos mentales reseñados al comienzo; en particular, llevar a un cambio conceptual.

Un experimento mental de la clase que nos interesa consiste en una situación contrafáctica caracterizada en base a los conceptos que están en cuestión. Para ser interesante, tal situación debe presentar algún rasgo paradójico, inesperado, incoherente. Lo cual sugiere que algo está mal en nuestra forma de entender el concepto, o bien que lo que es inadecuado es ese concepto y que debe ser modificado. Esto es lo distintivo de los experimentos que hemos destacado: que la situación misma sugiere tanto un diagnóstico del problema como el modo de evitarlo, por medio de un cambio conceptual.

En la terminología que hemos introducido en la sección

---

17 Este último ejemplo es de Peacocke, *op. cit.*

anterior, por tanto, podríamos decir que la clarificación de un concepto tiene lugar mediante la elaboración de una situación que sirve para explicitar la concepción implícita en que se sustenta, y poner de manifiesto sus implicaciones, en la medida en que sean problemáticas.

Por ello, la discusión de la situación contrafáctica exige valorar su posibilidad efectiva, la plausibilidad de la explicitación propuesta, la relevancia de otros aspectos no tenidos en consideración, etc. Pero si la anomalía es genuina es inevitable la reforma conceptual (por supuesto, puede haber casos menos "traumáticos" en los que puramente se rechace o modifique la concepción explicitada, y se avance hacia una mejor comprensión del concepto). De este modo se puede alcanzar un nuevo concepto sin recurrir a la abstracción a partir de la experiencia (como quiere el empirista) ni a una facultad de intuición metafísica (como quiere el racionalista). Para que ello sea posible, claro está, hacen falta otras capacidades además de la competencia semántica: de razonamiento hipotético, de mantenimiento de la consistencia, de construcción conceptual, etc. Sin entrar a fondo, quisiera mencionar la hipótesis, que se remonta a Mach, del origen adaptativo de estas capacidades, como base de su fiabilidad.

No obstante, podría parecer injustificada la conclusión de que el resultado de tales experimentos es un nuevo concepto –de inercia, de masa,...–. Creo que todo el mundo estaría de acuerdo en que el criterio de identidad conceptual no puede ser simplemente la homofonía del término con el que se expresa el concepto en el lenguaje. Sin embargo, dada la construcción en términos de conceptos y concepciones, ¿por qué no decir que lo que cambia es la concepción? Dado que la concepción no determina por sí misma el concepto, ¿en qué sentido el cambio en la concepción que se produce da lugar a un cambio conceptual, y no a una mera revisión de creencias?

La respuesta, me parece, es simple: en razón de los criterios externistas de la individuación conceptual. Ya Frege parte de que si las extensiones son diferentes, entonces los concep-

tos son diferentes. Y parece claro que las implicaciones a este nivel de los conceptos teóricos que hemos visto son claras. Se trata de formas distintas de concebir las "junturas de la naturaleza". De hecho, en este punto, las posibilidades son diversas. El resultado de un experimento mental de este tipo (o de cualquier otro medio para el cambio teórico) puede ser que un concepto resulte transformado en dos (el caso de Galileo que hemos presentado, con la distinción entre masa y peso), restrinja su campo de aplicación (lo que ocurrió con el concepto de gen), lo amplíe, o carezca de un único sustituto claro, sino que su referencia se divida entre dos o más. Este último caso es el que más difícil resulta para su comprensión, pues es el responsable de la aparición del fenómeno de la inconmensurabilidad. El ejemplo prototípico de este caso es el del concepto de masa newtoniana, que se escinde entre la masa inercial y la gravitatoria con la teoría especial de la relatividad.[18]

### *6. Experimentos mentales en filosofía*

Quisiera extender ahora este tratamiento de los experimentos mentales en la ciencia a los que se dan en filosofía. Sin pretender difuminar las diferencias entre los proyectos respectivos de la ciencia y la filosofía, creo que por lo que respecta a los experimentos mentales se basan en las mismas capacidades y su alcance, por tanto, es parecido. Dicho algo

[18] Ver Field (1973). Queda pendiente, desde luego, "el problema de Frege", esto es, la posibilidad de diferentes sentidos para un mismo referente. Me parece claro que este problema se acentúa para un externismo estricto, al individualizar los conceptos de un modo exclusivamente externista (el caso de Fodor, 1996, concluyendo con la esperanza que el problema de Frege no sea muy frecuente es sintomático). No puedo tratarlo aquí, pero considero dos posibles vías, complementarias, para su tratamiento. La primera consiste en notar que el problema de Frege parece depender del lenguaje; es decir, que se plantea sólo dado un medio representacional que permite diferentes signos para lo mismo; en segundo lugar, quizá podría tratarse como un caso de concepciones distintas de un mismo concepto. De todos modos, el problema de la individuación de los conceptos ("¿cuántos conceptos hay?) es distinto del problema de la posesión-competencia semántica ("¿qué conceptos tiene x?") que nos ocupa aquí, a pesar de su conexión obvia.

más abruptamente, el supuesto metodológico generalmente implícito en la práctica de la filosofía analítica[19] de que los experimentos mentales contribuyen al análisis conceptual, en el sentido clásico de la formulación de condiciones necesarias y suficientes, es erróneo.

La razón básica es que, si los experimentos mentales en filosofía se basan también en la competencia semántica/conceptual, no pueden tener el alcance metafísico que se pretende. Es decir, no pueden garantizar por sí mismos la posibilidad efectiva de ciertas situaciones, ni que puedan alcanzarse ciertas definiciones que constituyan verdades necesarias a priori.

En cambio, la comprensión de la metodología de los experimentos mentales dentro de la concepción analítica clásica de la filosofía, esto es, como análisis conceptual, como formulación de las condiciones necesarias y suficientes para que algo entre en la extensión de un concepto, pretende alcanzar este tipo de conclusiones. Un ejemplo de esta actitud, en relación a la filosofía de la mente, la expresa McGinn:

> "el filósofo trata de descubrir verdades necesarias a priori acerca [del fenómeno de la mente], verdades que se sostienen para cualquier ejemplificación posible del fenómeno mental en cuestión. Y tales verdades han de descubrirse precisamente mediante la elucidación del contenido de nuestros conceptos mentales."[20] (1982, p. 4)

Contenido al cual se accede por intuición, y que garantiza, *à la* Descartes, la coincidencia entre concepto y realidad. De lo que se trata, según este enfoque, es de imaginar situaciones concebibles donde se dé el concepto sin que sea claro que se dé alguna condición necesaria candidata para su determinación, o donde se pueda establecer la conexión necesa-

---

19 Como excepciones, por su explicitud, destacaría a McGinn *The Character of Mind*, Oxford, Oxford U.P., 1983; y Strawson G., *Mental Reality,* Cambridge, MS., MIT Press, 1994; La idea de que los experimentos mentales filosóficos también se basan en nuestra competencia semántica/conceptual se opone también por otra parte con quienes rechazan la relevancia de los experimentos mentales en filosofía, como Wilkes.

20 *Ibid*, p.4.

ria entre el concepto y la condición.

Tal como ocurre en la ciencia, la relevancia de un experimento mental depende de que plantee una posibilidad efectiva y problemática, por algún motivo, dados los conceptos disponibles, y que sugiera por tanto algún cambio al respecto. Sin embargo, esta compresión clásica parece dar por supuesto que la posibilidad metafísica deriva de la mera imaginabilidad, al modo en que de una contradicción lógica se deriva una imposibilidad metafísica. Un experimento mental probaría esa concebibilidad y, de este modo, su posibilidad. Sin embargo, por sí misma esa es una prueba muy débil. En cambio, a quien cuestiona esa posibilidad se le exige que demuestre que esa situación es metafísicamente imposible, no sólo físicamente imposible, sino autocontradictoria. O sea, un tipo de prueba muy exigente. Parece que hay una descompensación evidente, además de una dependencia de las nociones de analiticidad y necesidad metafísica. Es dudoso que haya tales necesidades metafísicas, más allá de lo nomológico. Pero sobre todo, es dudoso que la explicitación de nuestra competencia conceptual sirva para llegar a captarlas, esto es, que las podamos conocer a priori, sobre la base del análisis de nuestros conceptos. En primer lugar, porque, tal como la hemos presentado, la competencia semántica/conceptual no consiste en disponer de verdades necesarias, analíticas, que especifiquen la esencia del concepto. En segundo lugar, porque una cosa es lo que alguien considere como lo esencial de un concepto (a nivel epistemológico), y otra cosa es lo esencial de la extensión de ese concepto (a nivel ontológico).

En realidad, éstas son las lecciones que hemos aprendido con Quine, de su crítica a la distinción entre verdades analíticas y sintéticas, y de su defensa del holismo de la confirmación (otra cosa es el holismo semántico, para el que hace falta aceptar además el verificacionismo). Además, como hemos visto, el anti-individualismo semántico refuerza esta separación entre el nivel de la competencia y el ontológico, al tiempo que afirma la irreducibilidad de los conceptos a sus concepciones, y a sus posibles explicitaciones conscientes. Como

ya señalamos, la intuición central del anti–individualismo es que la ignorancia o el error con respecto a la naturaleza efectiva de algo no impide que seamos capaces de referirnos a ese algo. Dicho con Kripke, hay verdades necesarias a posteriori, por lo que hay que distinguir entre la fijación de la referencia y la identidad de la referencia. El enriquecimiento de la noción de poseer un concepto para que sea la base de la competencia semántica/conceptual que hemos propuesto no cuestiona ese punto central, sino que lo refuerza. En este contexto, lo que podríamos llamar el "test de existencia separada" del análisis conceptual, en base a la concebibilidad de una cosa sin la otra, carece de las garantías pretendidas, pues es posible imaginar situaciones "posibles" dada nuestra comprensión del concepto que no sean metafísicamente posibles (dada la naturaleza efectiva de las cosas a las que se refiere el concepto). Más bien refleja la influencia de la pretensión idealista de la racionalidad de lo real y la realidad de lo racional. En cambio, en realidad podemos imaginar imposibilidades sin reconocerlas como tales, del mismo modo que ciertas posibilidades pueden escapársenos.

Ello no quiere decir que los experimentos mentales elaborados dentro de ese marco metodológico no sean fructíferos. Es posible que muchos lo sean, pero según la propuesta que he desarrollado, no por las razones argüidas por los defensores del análisis conceptual, sino por la explicitación de la concepción asociada al concepto, y la puesta de manifiesto de su incompletud o inadecuación en algún sentido. Desde este punto de vista resulta más fácil explicar la etiología que presentan los experimentos mentales: la discusión en torno a la posibilidad efectiva de la situación imaginada, la oscilación entre el sentido habitual del concepto y el que emerge de un buen experimento mental, la sensación de éxito en la reformulación de un concepto, o la necesidad en ocasiones de la reforma conceptual, de ir más allá de los conceptos disponibles, es decir, de la dimensión constructiva de los experimentos mentales y su dependencia de la dimensión crítica.

En definitiva, aunque los objetivos de la ciencia y de la fi-

losofía difieren, en la medida en que ambas son actividades teórico–conceptuales, ambas están interesadas en la clarificación conceptual, y por tanto, ofrecen un ámbito de ejercitación a la metodología de los experimentos mentales, a la consideración de posibilidades contrafácticas. En la medida en que los experimentos mentales sirven para poner a prueba nuestros conceptos, al explicitar nuestras concepciones implícitas de ellos, pueden tener el mismo papel, y el mismo alcance, en la ciencia y en la filosofía.[21]

Universidad La Laguna
Dep. Hª y Fil. de la Ciencia
e-mail: agomila@ull.es

REFERENCIAS

BROWN, J.R., *The laboratory of the mind. Thought experiments in the natural sciences*, Londres, Routledge, 1991.

BURGE, T., "Wherein is language social?", En *Reflections on Chomsky*, A. George, Oxford, Blackwell.

BURGE, T., "Concepts, definitions and meaning", *Metaphilosophy* 24, (1989), 1993, pp. 309-325.

GIBBARD, A., *Wise choices, apt feelings,* Oxford, Clarendon Press, 1990.

GIERE, R. (ed.), *Cognitive models of science. Minnesota Studies in the Philosophy of Science*. Minneapolis: Minnesota U.P., 1992.

GOMILA, A. (en prensa, a), "Conceptos y concepciones: acerca del cambio conceptual", en *Palabras. Víctor Sánchez de Zavala In Memoriam*, K. Korta y F. García Murga (comp.), San Sebastián, Publicaciones Universidad del País Vasco.

GOMILA, A., "Thought experiments and semantic competence", *European Review of Philosophy* (en prensa).

21 Agradezco a Fernando Broncano el estímulo inicial para escribir este trabajo, y sus comentarios al núcleo de la propuesta. Versiones previas de este trabajo fueron presentadas en el *IV Meeting de la European Society of Philosophy and Psychology*, en Padua, así como en el *Encuentro del 25 aniversario de SADAF*, en Buenos Aires. Agradezco los comentarios de Josep Macià, Manuel García-Carpintero, Eduardo Rabossi, Diana Pérez, Tyler Burge y Susan Carey, a diversos aspectos del trabajo. Agradezco también el apoyo del Ministerio de Educación y Cultura, a través del proyecto PB95-0585.

GOODING, D., "The Procedural Turn", en Giere, 1992, pp. 45-76.
HEMPEL, C., *Aspects of Scientific Explanation*, New York, Free Press, 1965.
HIGGINBOTHAM, J., "How do I know what my words mean?", en *Knowing one's own mind*, Oxford, C. Wright & B. Smith, eds., 1997.
KOYRE, A., "Galileo's Treatise 'De motu gravium': the use and abuse of imaginary experiment", en *Metaphysics and Measurement*. Londres, Chapman and Hall, 1968.
KUHN, T., "A function for thought experiments", en *The Essential Tension*. Chicago, The University of Chicago Press, 1977.
MACH, E., "On thought experiments", en *Knowledge and Error*. Dordrecht, D. Reidel, (1905) 1976.
McGINN, C., *The Character of Mind*, Oxford, Oxford U.P., 1983.
NERSESSIAN, N., "How do scientists think? Capturing the dynamics of conceptual change in science", en Giere, 1992, pp. 3-44.
NORTON, J., "Thought experiments in Einstein's work", en *Thought experiments in Science and Philosophy*, T. Horowitz and G. Massey (eds.), Savage, MD Rowman and Littlefield, 1991.
PEACOCKE, C., "Implicit conceptions, understanding, and rationality", manuscrito, 1997.
SEARLE, J., *Intentionality*, Cambridge, Cambridge U.P., 1983.
SEARLE, J., *The rediscovery of the mind*. Cambridge, MIT Press, 1992.
SORENSEN, R., *Thought experiments*, Oxford, Oxford U.P., 1992.
STRAWSON, G., *Mental Reality,* Cambridge, MS., MIT Press, 1994.
WILKES, K., *Real People. Personal identity without thought experiments*. Oxford, Clarendon Press, 1988.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 89–105

VINCENZO P. LO MONACO

# PROBLEMAS CON LA SISTEMATICIDAD EN EL ANÁLISIS DE LA MENTE*

**Resumen.** En este trabajo me propongo examinar críticamente el dilema de Fodor acerca del carácter precario de una explicación conexionista de la cognición. Fodor, Pylyshyn y McLaughlin rechazan el conexionismo con el argumento de que sus partidarios no están en condiciones de explicar la sistematicidad sin recurrir a la implementación de una arquitectura clásica. Estimo que esta última afirmación descansa en un error. La conclusión del razonamiento tradicionalista sólo ha lugar si sus proponentes son capaces de formular una explicación neutral de la sistematicidad. Pero en ausencia de una explicación cabal -como pienso que es el caso- los conexionistas no confrontan ningún impedimento especial. En defensa de este punto de vista, muestro concretamente tres flancos débiles -i.e., la circularidad, la insuficiencia epistemológica y el atomismo reduccionista- que exhibe el argumento clásico de la sistematicidad. Finalmente, una vez establecido lo anterior, procedo a concluir que lo que, a mi juicio, está necesitado de explicación en la cuestión de la justificación de la representación cognitiva no es esencialmente la definición clásica de la sistematicidad, y que pueden haber otras modalidades de dar cuenta de la sistematicidad en el ámbito de la ciencia -por ejemplo, a la manera de la semántica holista de Davidson-, sin que por ello la representación conexionista quede excluida de estas formas legítimas de hacerlo.

**Palabras claves:** Teoría computacional de la mente, conexionismo, explicación atomista de la representación.

---

* La versión resumida de este artículo fue presentada como ponencia dentro del *Coloquio Semántica y Filosofía* en el XIV CONGRESO INTERAMERICANO DE FILOSOFIA tenido en la ciudad de Puebla, México, en agosto de 1999.

**Abstract:** In this paper, I examine and criticize Fodor's dilemma about the precariety of connectionist account of cognition. Fodor, Pylyshyn, and McLaughlin reject connectionism and argue that connectionists are unable to explain systematicity without implementing a classical architecture. I contend that this latter claim is based on a mistake: Traditionalist conclusion only seems to follow if they are able to sketch a neutral account of sistematicity. But in absence of a such fully explanation -as I think being the case-, connectionists bear no special burden in this matter. In support of this view, I set out three specific weaknesses -i.e., circularity, epistemological insufficiency, and atomism/reductionism- which affect the classical argument of systematicity. Finally, once that is done, I conclude that what seems to be crucial to explain regarding cognitive representation is not a classical definition of systematicity, and that there may be other ways to accommodate systematicity in scientific frame -say, in the manner of Davidson's holistic semantics- and indeed connectionist representation is one of these legitimate ways for doing that.

**Keywords:** Computational theory of mind, connectionism, atomistic account of representation.

En una ocasión Daniel Dennett dijo que "hablar de la mente es como hablar de sexo: ligeramente embarazoso, indecoroso y hasta deshonroso".[1] Aunque no dijo por qué, uno puede atreverse a conjeturar. Para el neurocientífico, hablar de la mente o "en mentalés" es simplemente deshonroso: el cerebro es una "máquina biológica" y todo lo que podemos decir de él es que responde al medio ambiente porque lo que entra como dato es procesado por actividades mecánicas gobernadas por reglas físicas. El psicólogo cognitivo es más cauto en sus apreciaciones. De lo que se trata –en su opinión– no es de perder la honra, sino de que parece más decoroso hablar de cerebro, hablar "en cerebral" subsumiendo el *mentalés*; al fin y al cabo no puede pasar por alto que lo que se está estudiando es la comprensión de un sistema intencional. Los filósofos suelen ser, en general, mucho más liberales, tal vez por el hecho de estar hablando "en mentalés" desde hace más de dos milenios. Para éstos, si algo hay de embarazoso, se trata de un embarazo ligero y fútil, que perdura tan sólo hasta elegir si hablar de la mente "en mentalés" o hacerlo "en

1 Cf. Dennett, D., *La actitud intencional*, Barcelona, Gedisa, p. 15.

cerebral", o bien en algún otro lenguaje, como por ejemplo el de la "superviniencia"

La inteligencia artificial (IA) parecía en sus inicios, en la década de los cincuenta, seguir el estilo *ouvert* de los filósofos; aunque sus primeros modelos dependieron decididamente de los avances técnicos de las supercomputadoras modernas, no desdeñaban el empleo de las metáforas biológicas y psicológicas más usuales en la descripción de los sistemas intencionales, hasta el punto de proclamar, a la sazón, que "...La intuición, la comprensión y el aprendizaje no son ya posesión exclusiva de los humanos: cualquier computadora grande de alta velocidad también puede ser programada para exhibirlos".[2] Pero muy pronto tal liberalidad se trocó en intolerancia. A medida que crecía la confianza en el procesamiento simbólico de la información como modelo de la teoría computacional de la mente, aumentaba también la hostilidad hacia cualquier otra perspectiva de análisis que no partiese exclusivamente de la representación simbólica. El primero en experimentar en carne propia los rigores de tal intolerancia fue Frank Rosenblatt, pionero de la modelización computacional con redes neuronales. La forma en que Rumelhart y McClelland[3] narran el despiadado ataque que Minsky y Papert emprendieran contra la incipiente perspectiva conexionista, merece ciertamente figurar como episodio admirable de lo que Lakatos llamara "historia externa" de la ciencia. Para 1970, en lo que concierne a la investigación en IA, los fondos gubernamentales habían vuelto a los bolsillos de los representacionalistas simbólicos del M.I.T. y, como lo recuerda Newell,[4] el perceptrón de Rosenblatt no era más que un recuerdo lejano.

Sin embargo, durante la década pasada la aproximación conexionista ha reaparecido con nuevo vigor en el panorama computacional del análisis de la mente, logrando en pocos años notables avances en la tarea de diseñar la actividad computacional como la función de una red conexionista que reemplaza los símbolos y las reglas de símbolos con patrones de actividad numérica sobre grupos de unidades y patrones de pesos distribuidos en paralelo. Recuperando la idea de

---

2 Simon, H. y Newell, A., ¨Heuristic Problem Solving: The Next Advance in Operations Research¨. *Operations Research* 9 (1958), p. 6.

3 Cf. Rumelhart, D.E. y McClelland, J.L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. I, Cambridge, MIT Press, 1986, p. 111 y ss.

4 Cf. Newell, A., ¨Intellectual Issues in the History of the Artificial Intelligence¨, en Machlup, F. y Mansfield, U. (eds.), *The Study of Information: Interdisciplinary Messages*, Nueva York, Wiley, 1983, pp. 10-12.

Rosenblatt, estas redes son concebidas en analogía con modelos neurales simplificados y su plausibilidad depende de la hipótesis de que el papel de los símbolos y reglas en el computacionalismo clásico es ahora desempeñado por algo semejante a las neuronas y las sinapsis. De ahí que los sistemas conexionistas constituyan hoy una explicación computacional de los procesos neurales; de ahí también que el conexionismo haya hoy resurgido como un rival no desestimable de la aproximación que simula la cognición a través de procesos simbólicos, en la medida en que se ha reconocido que el proceso distribuido en paralelo de los modelos conexionistas los hace idóneos para ejecutar determinadas tareas –por ejemplo, los patrones de reconocimiento, la categorización y el pensamiento analógico– que resulta muy difícil realizar en los procesos simbólicos clásicos.

A pesar de los recientes resultados obtenidos por el conexionismo en la tarea de simular con algún detalle los procesos que las personas emplean para solucionar problemas diversos, y en general en la tarea de construir un modelo plausible de la cognición humana, existe todavía gran suspicacia en torno a sus virtudes explicativas y abundan sus detractores, especialmente entre los filósofos. Ya no se la denigra con los argumentos de la "explosión combinatoria" o la "propagación hacia atrás de los errores", que habían mandado a las redes neuronales de Rosenblatt a descansar momentáneamente.[5] Ahora se arguye que hay serios problemas con la sistematicidad en los modelos conexionistas. Concretamente, lo que algunos filósofos –como Fodor y Pylyshyn,[6] McLaughlin[7] y diversos otros– encuentran desde el inicio confuso en la perspectiva conexionista es la pretensión de

---

5 Cf. Rumelhart y McClelland, *Parallel Distributed Processing...*, cit., p. 32 y ss.

6 Cf. Fodor, J.A. y Pylyshyn, Z.W., "Connectionist and the cognitive architecture: A critical analysis", *Cognition* 28 (1988), pp. 3-71.

7 Cf. Fodor, J.A. y McLaughlin, B.P., "Connectionist and the problem of systematicity: Why Smolensky's solution doesn't work?", *Cognition* 35 (1990), pp. 183-204. Véase también MacLaughlin, B.P., "The Connectionism/Classicism Battle to Win Souls", *Philosophical Studies* 71 (1993), pp. 163-190.

atribuir al proceso distribuido en paralelo la capacidad de proporcionar una explicación de la representación de la cognición. El punto fundamental de la mayor parte de las críticas que apuntan a una objeción inmediata y concluyente contra aquella pretensión es, para parafrasearlo muy esquemáticamente, el siguiente: según la teoría computacional de la mente, los sistemas tienen estados mentales debido a que implantan las representaciones en códigos y las disponen según una relación particular. En esta teoría de la representación, la cognición como actividad es un proceso de operaciones formales que tienen lugar en representaciones sintácticamente estructuradas.[8] Muy diversamente, dado que el conexionismo carece formalmente de modularidad simbólica,[9] sus modelos no están capacitados para proveer estados mentales o cognitivos, pues adolecen de una estructura sintáctica que relacione las representaciones. Y este es justamente el punto de Fodor:

> ...Una teoría cognitiva adecuada desde el punto de vista empírico debe reconocer no sólo relaciones causales entre los estados representacionales, sino también relaciones de constitución de tipo sintáctico y semántico; por ende, [...] la mente no puede ser, en su estructura general, una red conexionista".[10]

Ahora bien, me parece que esta última afirmación descansa en un error. Creo que la conclusión tradicionalista sólo parece seguirse debido a la confusión que ha caracterizado al debate, en la medida en que lo que los filósofos entienden normalmente por 'representación semántica' ha sido manipulado y convertido en algo trivial. Sostengo que la teoría computacional clásica es semánticamente referencialista y atomista en el sentido filosófico ordinario, y que el atomismo se encuentra en serios problemas a la hora de explicar cómo el concepto en la cabeza llega a ser el tipo de cosa que puede representar al mundo. Mantengo, además, que debido al

---

8 Cf. van Gelder, T., "Compositionality: A Connectionist variation on a classical theme", *Cognitive Sciences* 14 (1990), pp. 355-384.

9 Cf. Bechtel, W., "Current connectionism", *Minds and Machines* 3 (1993).

10 Fodor y Pylyshyn, "Connectionist and the cognitive...", cit., p. 32 [La traducción es nuestra].

hecho de que el conexionismo propone un modo de representar los contenidos semánticos significativamente diferente del análisis composicional de la cognición,[11] los conexionistas están legítimamente facultados para cambiar lo que vale como reconocimiento de patrones de constitución sintáctica y semántica.[12]

Comenzaré por referirme a lo que he llamado en otra parte[13] el "dilema de Fodor". En efecto, Fodor ha sostenido que hay un dilema que, en su opinión, afecta al conexionismo al colocarlo en una situación paradójica. El dilema es como sigue:

> ... si el conexionismo no puede dar cuenta de la sistematicidad, falla entonces en proporcionar una base adecuada para una teoría de la cognición; pero si su explicación de la sistematicidad requiere de procesos mentales que son sensitivos a la estructura constituyente de las representaciones mentales, entonces la teoría de la cognición que ofrece será, en el mejor de los casos, la implementación de la arquitectura de un modelo 'clásico' (de lenguaje del pensamiento) [...].[14]

Como Fodor lo caracteriza ahí, el dilema surge al explicar la sistematicidad en una teoría de la cognición. Pero la paradoja que engendra -y que afectaría al conexionismo- tiene implicaciones para explanar los poderes representacionales en una explicación semántica de los estados cognitivos. En lo sucesivo deseo examinar estas implicaciones en relación con la imposibilidad de una explicación conexionista de la representación semántica. Por simplicidad, los dos cuernos del dilema se explorarán de manera separada, a fin de mostrar lo que hay de problemático en la crítica de Fodor y McLaughlin. En lo que concierne al segundo cuerno del dilema, he mos-

---

11 Cf. Smolensky, P., "Tensor product variable binding and the representation of symbolic structures in connectionist systems", *Artificial Intelligence* 46 (1990), pp. 159-216. También van Gelder, "Compositionality..., cit., p. 364 y ss.

12 Véase sobre este punto Lo Monaco, V.P., "The Computational Theory of Mind and Searle's Problem", en Callaos, N. (ed.), *Proceedings of ISAS '96*, Orlando, IIIS, 1996, pp. 328-335. También Smolensky, P., Legendre, G. y Miyata, Y., *Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition,* Cambridge, M.I.T. Press, 1994.

13 Cf. Lo Monaco, V.P., "Connectionism, Systematicity, and Fodor´s Dilemma", en Callaos, N. y Martin, G. (Eds.), *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, vol. IV, Caracas, IIIS, pp.83-88.

14 Fodor y Pylyshyn, "Connectionist and the cognitive...", cit., p. 9 [La traducción es nuestra].

trado ya en otra parte[15] que el argumento de la implementación es incoherente, porque los modelos semánticos conexionistas son holísticos, y en consecuencia podemos afirmar que, dentro de un marco conexionista concebido, por ejemplo, en los términos de la teoría del lenguaje de Davidson, no existe dificultad ninguna para explicar intrínsecamente la representación semántica sin implementar una arquitectura clásica.[16] Pasemos entonces directamente al primer cuerno del dilema de Fodor –i.e., la afirmación de que el conexionismo no puede dar cuenta de la sistematicidad.

Tradicionalmente, a diferencia de la mayor parte de los conexionistas, los computacionalistas clásicos hacen un énfasis desmedido en la sistematicidad para explicar la estructura constituyente de los estados representacionales.[17] Sus razones comprenden usualmente el carácter *compositivo* de la representación. Arguyen que los estados cognitivos "heredan" sus propiedades semánticas y su intencionalidad *de manos* de las propiedades semánticas de las representaciones mentales, de modo tal que éstas deben tener una estructura interna compuesta de partes elementales –i.e., constituyentes. Para explicar esta propiedad representacional de los estados cognitivos como la capacidad que poseen los estados mentales de relacionar sus contenidos en determinados modos sistemáticos, proponen una construcción computacional de los estados intencionales en términos de la arquitectura clásica conocida como "Teoría Representacional de la Mente".[18] Ésta construye los contenidos semánticos echando mano de un código de un lenguaje simbólico para representar objetos, relaciones, eventos, acciones, etc., y empleando dispositivos "de entrada" para formar representaciones simbólicas, las cuales son en-

---

15 Cf. Lo Monaco, "Connectionism...", cit., pp. 84-86.

16 Cf. Smolensky, P., "Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture", en McDonald, C.G. y McDonald, G. (eds.), *The philosophy of Psychology: Debates on Psychological Explanation,* Oxford, Blackwell, 1994.

17 Cf. Bradshaw, D.E., "Connectionism and the Specter of Representationalism", en Horgan, T. y Tienson, J. (eds.), *Connectionism and the Philosophy of Mind*, Dordrecht, Kluwer, 1991. También van Gelder, "Compositionality...", cit., pp. 360-366.

17 Cf. Fodor, J.A., *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MIT Press, 1987. Del mismo autor, *A Theory of Content and Other Essays*, Cambridge, MIT Press, 1990.

tonces relacionadas con una estructura semántica, de tal modo que exista una correspondencia término a término entre símbolos y contenidos semánticos, donde cada símbolo representa un concepto definido. De ahí que, si se toma en cuenta que las manipulaciones simbólicas son regidas por reglas de inferencia válidas, entonces las interacciones de símbolos y contenidos semánticos preservarán los valores de verdad y el marco semántico resultará definido causalmente de un modo tal que el significado de un símbolo consistirá en "... el conjunto de cambios que éste (*el símbolo*) provoca que el sistema efectúe, o como entrada o bien como respuesta a algún estado (interno o externo)". En definitiva, puede decirse que la arquitectura clásica ejecuta sus operaciones sobre representaciones simbólicas a guisa de una sintaxis composicional que presupone una semántica composicional.[19]

Ahora bien, ¿rige en esta arquitectura el concepto de sistematicidad? En verdad, una respuesta definitiva a esta cuestión es realmente controversial. Fodor, Pylyshyn y McLaughlin han sugerido al menos tres formas de definir tentativamente la sistematicidad:

(i) la propiedad de tener capacidades cognitivas relacionadas para estados intencionales,[20] o

(ii) la propiedad de tener capacidades cognitivas relacionadas para estados intencionales cuya posesión implica la posesión de 'bases constitutivas' en virtud de las cuales quien típicamente posea una de estas capacidades posee también la otra (u otras),[21] o

(iii) la posesión de determinadas capacidades sustantivas (*"molar"*) cuyas condiciones de satisfacción constituyen la posesión de determinadas capacidades constitutivas y de sus interacciones.[22]

Hay una larga tradición en filosofía que intenta articular

---

19 Cf. Bradshaw, "Connectionism and the Specter...", cit., pp. 67-70; van Gelder, "Compositionality...", cit., pp. 363-369.

20 Cf. Fodor y Pylyshyn, "Connectionist and the cognitive...", cit., p. 38.

21 Cf. Fodor y McLaughlin, "Connectionist and the problem...", cit., pp. 186-187; MacLaughlin, "The Connectionism/Classicism Battle...", cit., p. 168.

22 Cf. MacLaughlin, "The Connectionism/Classicism Battle...", cit., pp. 170-172.

los lineamientos esenciales de una lógica de la definición. Dado el alcance teórico de la definición clásica de la sistematicidad, me dispongo a examinarla a la luz de aquellos lineamientos, pues tengo la sospecha de que la definición de marras es o bien trivial o, en el mejor de los casos, dependiente del contexto. Para explicar el punto que pretendo establecer –esto es, que el argumento de la sistematicidad levantado contra el conexionismo es demasiado débil–, es importante hacer dos aclaraciones. Primero, es evidente que hay algunas diferencias importantes entre las tres definiciones de la sistematicidad referidas. Por supuesto, la definición (i) es simple y básica, mientras que las definiciones (ii) y (iii) añaden, además, los requerimientos de 'constitución' y 'satisfacción', progresiva y respectivamente. De ahí que, segunda aclaratoria, aunque es contundente pensar que las definiciones expresan lo mismo, porque la definición (i) habría podido ser complementada de muchos otros modos distintos de lo que afirman las definiciones (ii) y (iii), parece razonable asumir que, al menos a los efectos de la crítica al conexionismo, la definición (iii) las incluye a todas y, acorde con los computacionalistas clásicos, es capaz de caracterizar a la sistematicidad de manera cabal y satisfactoria.

Sin perder de vista estas clarificaciones, permítaseme ahora regresar a mi hipótesis acerca de la debilidad del argumento clásico de la sistematicidad. En su defensa, ofrezco las siguientes razones.

Una primera razón atañe a lo que es un requisito *sine qua non* de la teoría de la definición, a saber, la no circularidad. Una explanación de la sistematicidad basada en la posesión de determinadas capacidades substantivas, entendidas como capacidades constitutivas que resultan satisfechas por su posesión, podría ser circular si la justificación de las bases constitutivas se identificara automáticamente con la posesión de capacidades cognitivas. Considérese el caso de las emisiones lingüísticas, donde la relación de sistematicidad es mediada por los estados cognitivos de hablantes y oyentes. En este ámbito, determinadas capacidades cognitivas podrían ser

sistemáticamente relacionadas si, y sólo si, *como cuestión de necesidad nomológica*,[23] un parlante de un lenguaje L que comprende una L-oración dada, comprende también otra L-oración relacionada con aquélla. El problema de cómo un L-parlante puede comprender por necesidad nomológica L-oraciones relacionadas, es suplantado entonces por un problema acerca de las capacidades cognitivas en el "lenguaje del pensamiento". La Teoría Representacional de la Mente proporcionaría un modelo para explicar la sistematicidad por esa vía. He aquí algunos detalles de la paráfrasis: dado un parlante S y un par de L-oraciones, P y Q (por ejemplo, "Juan ama a la chica" y "La chica ama a Juan", respectivamente), puesto que *significar (que) P* entraña un estado mental A de S y *significar (que) Q* entraña un estado mental B de S, existe una representación estructurada RP↔RQ tal que S tiene A si, y sólo si, S tiene Q, y a la inversa, pues se supone que S posee la misma capacidad cognitiva constitutiva que construye al unísono tanto A como B.[24] Pero si ésta es la caracterización, entonces la estrategia para explicar la sistematicidad es realmente circular. Ella se hace descansar en el hecho que las capacidades cognitivas resultan explicadas por recurso a la interacción de significados de las oraciones, pero la interacción de significados de las oraciones es explicada a su vez por recurso a los estados mentales implícitos en la asignación de intencionalidades relacionadas a la misma capacidad cognitiva. De resultas, pareciera que el punto crucial a explicar no es la relación entre oraciones, representaciones semánticas y determinados estados y procesos cognitivos, sino más bien el carácter *constitutivo* de las capacidades cognitivas.

Prescindiendo de la circularidad, la explicación clásica de la sistematicidad presenta otra gran debilidad. Ésta es sugerida por el análisis del concepto de sistematicidad de Fodor que lleva a cabo R. Matthews.[25] En el análisis de Matthews, la ra-

---

23 Cf. *ibid.*, p. 167.

24 Cf. Fodor, *Psychosemantics...*, cit., pp. 99-108.

25 Cf. Matthews, R.J., "Three-Concept Monte: Explanation, Implementation and Systematicity", *Synthese* 101 (1994), pp. 347-63.

zón por la cual la sistematicidad resulta supuestamente un argumento intranscendente es que "...no existen explicaciones clásicas de la sistematicidad que merezcan ese nombre, en particular en el constructo de la sistematicidad de Fodor y otros como la propiedad de tener capacidades cognitivas sistemáticamente relacionadas para los estados intencionales".[26] Aunque se ha sostenido que la sistematicidad puede explicarse en términos funcionales por recurso a la Teoría Representacional de la Mente, la razón por la cual Matthews insiste en que la sistematicidad es epistemológicamente intranscendente es que los computacionalistas clásicos no están actualmente en condiciones de proporcionar una sintaxis composicional y una psicosemántica naturalista.[27] Esta última aseveración encuentra un sostén más preciso en filósofos como Loewer y Rey,[28] pero la idea básica es la siguiente: la mayor parte de la investigación que se realiza en psicosemántica en perspectiva computacional en relación con la explicación del significado de oraciones, encuentra serios tropiezos a la hora de mostrar que sus modelos proporcionan algo semejante a una explicación exitosa de la cognición con base en una sintaxis composicional. Matthews establece este punto insistiendo en que lo que los clásicos describen como explicaciones de la sistematicidad son en realidad meras estipulaciones, "...algunas ideas muy generales de cómo podría construirse tal explicación".[29] Mantengo el mismo punto al insistir no sólo en que los modelos computacionales clásicos nada tienen en el ámbito conceptual que pueda contar como explicación de las capacidades cognitivas e indique a la vez sistematicidad, sino también en que hay algo que permanece normalmente oculto en la discusión generada en torno al debate conexionismo/clasicismo –una discusión en la cual el desafío de Fodor de explicar la racionalidad está cayendo velozmente en una mera confusión-, a saber, la naturaleza problemática de las

---

26 *Ibid.,* p. 356 [La traducción es nuestra].
27 Cf. *Ibid.* p. 358.
28 Cf. Rey, G. y Loewer, B. (eds.), *Meaning in Mind: Fodor and Its Critics*, Cambridge, Blackwell, 1991.
29 Matthews, "Three-Concept Monte...", cit., p. 360.

categorías de "significado" y 'referencia',[30] y esta situación nos plantea a su vez un desafío que trasciende el escenario de una teoría empírica de la sistematicidad tal como Fodor la imagina. Concebida en estos términos, la cuestión nos arrastra más lejos, al punto de señalar una nueva debilidad.

Para que se comprenda a plenitud esta tercera debilidad en el argumento de la sistematicidad, necesito ahora referirme a la cuestión de cómo un lenguaje del pensamiento explica lo que es representar para un lenguaje natural en términos de lo que sus oraciones realmente representan. Un modo de hacer esto es ilustrado por Fodor y Lepore[31] argumentando a favor de una explicación atomista de las representaciones mentales como la única forma de establecer las condiciones de identidad de los contenidos mentales sin describir el estado mental total del organismo. Empero, dado que se explica la verdad de una oración en términos de la referencia de sus partes, la explicación composicional de la representación está necesitada de una explicación reductiva de la referencia de los términos que constituyen una oración. De ahí que el análisis composicional clásico sea atomista en un doble sentido. Es, en primer lugar, atomista porque explica lo que es para un estado mental representar un objeto por recurso a relaciones causales entre el mundo y la mente, donde los *relata* son análogos mentales de las palabras y partes del mundo. En segundo lugar, es atomista en tanto asocia los análogos de las palabras mentales con sus extensiones por el procedimiento de correlacionar, por un lado, palabras mentales y objetos, y por el otro predicados mentales y propiedades. Pero es además reductivo, pues emplea los recursos disponibles para explicar la verdad de una oración en términos de la referencia de sus componentes más simples.[32]

---

30 Cf. Bradshaw, "Connectionism and the Specter...", cit., pp. 69-73; también Horgan, T. y Tienson, J., "Settling into New Paradigm", *Southern Journal of Philosophy* 26 (1987), pp. 97-113.

31 Cf. Fodor, J.A. y Lepore, E., *Holism: A shopper's guide*, Cambridge, Blackwell, 1992.

32 Cf. Hogan, M., "What is Wrong with an Atomistic account of Mental Representation", *Synthese* 100 (1994), pp. 307-327.

No obstante, todo esto no parece haber inquietado en demasía a Fodor, pues en escritos todavía recientes[33] no ha tenido empacho alguno en conceder que la explicación composicional de la representación que propone es atomista. En defensa del atomismo semántico, entendido como "...la idea de que lo que se significa es enteramente independiente de lo que se cree...",[34] ha afirmado que esa es la única alternativa al holismo.[35]

Por mucho que esta afirmación pueda resultar verdadera, tiende, no obstante, a ocultar el hecho que el atomismo presenta dificultades por su propia cuenta. Una huella de tales dificultades puede encontrarse en la descripción que hace Davidson de cómo dar cuenta del significado en un lenguaje natural en términos suficientes para comprender las emisiones de los hablantes de ese lenguaje.[36] Ahí Davidson concibe el atomismo como 'el método del bloque constructivo', aquel que "...empieza con lo simple y construye hacia arriba...", y lo rechaza porque pretende "...dar una caracterización no lingüística de la referencia, pero no parece haber chances de ello".[37] Dice que "no parece haber chances" porque está asumiendo, siguiendo a Quine, que "...la totalidad de la evidencia disponible para un oyente no determina una forma única de traducir las palabras de un hombre en las de otro[...]".[38] Una vez abrazada la tesis de la 'inescrutabilidad de la referencia', Davidson concluye que la referencia misma "...quedará en el camino. Ella no desempeña una función esencial en la explicación de la relación entre lenguaje y realidad".[39] Pienso que Davidson está en lo cierto cuando afirma que el atomismo es una postura indefendible, pero creo además que el argumento que él ofrece en su contra se queda

---

33 Cf. Fodor, *A Theory of Content...*, cit., pp. 17-23; Fodor y Lepore, *Holism...*, cit., pp. 44-45.
34 Fodor, *A Theory of Content...*, cit., p. 21.
35 Cf. Fodor y Lepore, *Holism...*, cit., pp. 47
36 Davidson, D., "Reality without reference", en Platts, M. (ed.), *Reference, Truth and Reality,* 1980, pp. 131-140.
37 *Ibid.*, p. 135.
38 *Ibid.*, p. 137.
39 *Ibid.*, p. 140.

corto, en especial si se conecta el atomismo con el *representacionalismo*. De hecho, cuando se lo combina con una aproximación semántica representacional, el atomismo es insostenible a causa de tres importantes consideraciones.

En primer lugar, el atomismo representacional da por sentado que existen los estados mentales (las representaciones), los cuales en algún sentido contienen partes significantes (conceptos) que son establecidas sólo por inferencia lógica a partir de la estructura del lenguaje. Dado que la evidencia necesaria para establecer la existencia de estados mentales simples (conceptos), considerados en el modelo atomista como significados de las palabras, depende de la inferencia a partir del lenguaje, y puesto que todas las inferencias lógicas a partir del lenguaje presuponen únicamente una caracterización completa de los rasgos semánticos de sus partes (las oraciones), aunada a las relaciones formales que rigen las oraciones T de Tarski como vehículos para dar indirectamente un contenido empírico a las relaciones entre nombres -o predicados- y objetos, concluimos que la existencia de representaciones mentales como correlatos psicológicos de oraciones no puede ser establecida con algún grado de certidumbre. La razón de ello es que la evidencia obtenida de las relaciones lingüísticas es sólo y siempre *evidencia lingüística*, en cuyo caso la cuestión de cómo hacer uso del contenido mental que una oración presuntamente entraña como relación entre una representación mental y un hecho extralingüístico, queda sin respuesta.

La segunda consideración es que si se admite que los significados de palabras son 'objetos mentales' o 'conceptos en la cabeza' susceptibles de construcción, de retención y de evocación a través de la mente, en el supuesto de que sean inmanentes a los actos intencionales y cognitivos de la mente misma, la postulación de contenidos mentales como significados correctamente identificables de las palabras está necesitada de una historia plausible de cómo los valores semánticos de los conceptos simples se combinan unos con otros en la mente para finalmente formar los valores semánticos de las repre-

sentaciones mentales.

Por último, tercera consideración, el ardid de igualar los significados de las palabras (nombres y predicados) y los objetos mentales presupone que las palabras que se usan en el lenguaje están causalmente conectadas con sus análogos mentales por medio de cierta técnica. El primer caso constituye esencialmente un recurso al lenguaje del pensamiento para explicar y validar la diferencia entre tener creencias y tener meramente una ristra de conceptos en la mente. El segundo, es un recurso a una caracterización reductiva de la referencia de los términos que constituyen una oración, en correspondencia con una esquematización funcional del significado de los conceptos que constituyen una representación. Para resumir, en ambos casos se trata de cuestiones que el atomista representacional tiene dificultad en explicar. La primera se autoaniquila, pues la Teoría Representacional de la Mente está aún lejos de proporcionar un modo satisfactorio de explicar cómo se combinan los valores semánticos de las representaciones mentales sin sencillamente presuponer la combinación misma. La segunda es realmente una cuestión problemática, si se tiene en cuenta que el atomista representacional no dispone de una explicación causal que permita construir una función vicaria que cubra la trayectoria que va de los objetos/valores de verdad a los conceptos/representaciones mentales, y a la inversa. En consecuencia, si en una teoría del lenguaje que proporciona las condiciones bajo las cuales todas sus oraciones representan fielmente el mundo, no hay nada que indique la necesidad de recurrir a estados mentales representacionales para fijar la relación lenguaje-mundo, entonces una explicación atomista de la representación puede tornarse del todo irrelevante.

En conclusión, en la perspectiva desde la cual hemos enfocado el asunto, el argumento de la sistematicidad se ha vuelto intrascendente. Irónicamente, el desafío de explicar la sistematicidad que los tradicionalistas proponen a los conexionistas no puede ser acometido por aquéllos, pues los computacionalistas clásicos adolecen de problemas más gra-

ves que los que ellos mismos critican en otras formas de explicar la cognición. ¿Qué ocurre entonces con la sistematicidad? Como ya hemos indicado, cabe imaginar otras formas de dar cuenta de la sistematicidad y aún rechazar, no obstante, el dilema de Fodor, pero no ha sido mi propósito aquí explorar esas posibilidades. Mi interés se ha centrado exclusivamente en un punto específico, a saber: el desafío de Fodor de explicar la sistematicidad no atañe al conexionismo más de lo que atañe al computacionalismo clásico mismo. En realidad, el debate entre clasicistas y conexionistas apenas empieza. Creo que en ambas posiciones se han dado promisorios avances que constituyen desarrollos no desestimables, pero mucho queda aún por hacer en el nivel de la explicación científica. Puede que los argumentos filosóficos sean deseables para indicar dónde falla la teoría, pero es la ciencia la que en definitiva hace el trabajo.

Instituto de Filosofía
Universidad Central de Venezuela

## REFERENCIAS

Bechtel, W., "Current connectionism", *Minds and Machines* 3, (1993).
Bradshaw, D.E., "Connectionism and the Specter of Representationalism", en Horgan, T. y Tienson, J. (eds.), Connectionism and The Philosophy of Mind, Dordrecht, Kluver, 1991.
Clark, A., "Special issue: Philosophical issues in connectionist modeling", *Connection Science* 4, (1992), pp. 171-173.
Davidson, D., "Reality without reference", en Platts, M. (ed.), *Reference, Truth and Reality,* 1980, pp. 131-140.
Fodor, J.A., *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MIT Press, 1987.
Fodor, J.A., *A Theory of Content and Other Essays*, Cambridge, MIT Press, 1990.
Fodor, J.A. y Lepore, E., *Holism: A shopper's guide*, Cambridge, Blackwell, 1992.
Fodor, J.A. y McLaughlin, B.P., "Connectionist and the problem of systematicity: Why Smolensky's solution doesn't work?", *Cognition* 35, (1990), pp. 183-204.
Fodor, J.A. y Pylyshyn, Z.W., "Connectionist and the cognitive architecture: A critical analysis", *Cognition* 28, (1988), pp. 3-71.

Hogan, M., "What is Wrong with an Atomistic account of Mental Representation", *Synthese* 100 (1994), pp. 307-327.
Horgan, T. y Tienson, J., "Settling into New Paradigm", *Southern Journal of Philosophy* 26 (1987), pp. 97-113.
Horgan, T. y Tienson, J(eds.), *Connectionism and the Philosophy of Mind*, Dordrecht, Kluwer, 1991.
Lo Monaco, V.P., "The Computational Theory of Mind and Searle's Problem", en Callaos, N. (ed.), *Proceedings of ISAS '96*, Orlando, IIIS, 1996.
Lo Monaco, V.P., "Connectionism, Systematicity, and Fodor´s Dilemma", en Callaos, N. y Martin, G. (Eds.), *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, vol. IV, Caracas, IIIS, 1997.
MacLaughlin, B.P., "The Connectionism/Classicism Battle to Win Souls", *Philosophical Studies,* 71 (1993), pp. 163-190.
Matthews,R.J., "Three-Concept Monte: Explanation, Implementation and Systematicity", *Synthese* 101 (1994), pp. 347-63.
Newell, A., "Intellectual Issues in the History of the Artificial Intelligence", en Machlup, F. y Mansfield, U. (eds.), *The Study of Information: Interdisciplinary Messages*, Nueva York, Wiley, 1983
Rey, G. y Loewer, B. (eds.), *Meaning in Mind: Fodor and Its Critics*, Cambridge, Blackwell, 1991.
Rumelhart, D.E. y McClelland, J.L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. I, Cambridge, MIT Press, 1986.
Simon, H. y Newell, A., "Heuristic Problem Solving: The Next Advance in Operations Research". *Operations Research* 9 (1958).
Smolensky, P., "Tensor product variable binding and the representation of symbolic structures in connectionist systems", *Artificial Intelligence* 46 (1990), pp. 159-216.
Smolensky, P., "Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture", en McDonald, C.G. y McDonald, G. (eds.), *The philosophy of Psychology: Debates on Psychological Explanation,* Oxford, Blackwell, 1994.
Smolensky, P., Legendre, G. y Miyata, Y., *Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition,* Cambridge, MIT Press, 1994.
Van Gelder, G., "Compositionality: A Connectionist variation on a classical theme", *Cognitive Sciences* 14 (1990), pp. 355-84.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 107–126

LEVIS ZERPA MORLOY

# FUNDAMENTOS LÓGICOS DE LAS REDES NEURALES ARTIFICIALES[1]

**Resumen:** En este trabajo exponemos en forma parcial una reconstrucción lógica, basada en la concepción no–enunciativa o estructuralista de las teorías científicas, de las redes neurales artificiales (deterministas) de una y dos capas, o más específicamente, del perceptrón de una y dos capas. Definimos los modelos de la teoría general y también una versión más restringida, el perceptrón binario y su interpretación geométrica. También analizamos brevemente un conocido problema lógico relacionado con estas redes: la representación de la disyunción exclusiva (*"XOR problem"*) y de las demás funciones no linealmente separables. Examinamos una solución concatenando redes de una capa para obtener redes multicapa. Esto amplía el poder predictivo de la red y presenta interesantes aspectos metodológicos.

**Palabras Claves:** Inteligencia artificial conexionista, metateoría estructuralista, red perceptrón.

**Abstract:** In this paper we develop, in a partial way, a logical reconstruction of the (deterministic) artificial neural nets of one and two layers, the perceptron, based in the no-statement or structuralist philosophy of science. We define the models of the general theory and we define, too, a more restricted version of it, and of the binary perceptron and its geometrical interpretation. A well-known problem of those nets are analyzed: the representation of the exclusive OR ("XOR problem") and the others no linearly separable functions. We consider a solution of this problem joining

[1] Una primera versión de este trabajo fue leída como ponencia en el V Congreso Nacional de Filosofía (Caracas, noviembre de 1999). Agradecemos al Prof. José Burgos por su valiosa colaboración en el desarrollo de la investigación previa a este trabajo. También agradecemos al Prof. Ricardo Chang por su colaboración en esa etapa.

or concatenating one-layer nets to obtain multilayer nets. This enhance the predictive power of the net and is a source of interesting methodological aspects.

**Keywords:** Connectionist artificial intelligence, structuralist metatheory, perceptron network.

## *§ 1 Introducción*

En este trabajo exponemos en forma parcial una reconstrucción lógica, basada en la concepción no–enunciativa o estructuralista de las teorías científicas[2], de las redes neurales artificiales (deterministas) de una y dos capas, o más específicamente, del perceptrón de una y dos capas[3]. También analizamos brevemente un conocido problema lógico relacionado con estas redes: la representación de la disyunción exclusiva (*"XOR problem"*) y de las demás funciones no linealmente separables.

Hay que comenzar por decir que el tema de las vinculaciones entre la Inteligencia Artificial (= IA) y la Lógica Matemática se trata de manera muy distinta en la IA Simbólica y la IA Conexionista. En la IA Simbólica se reconoce el destacado papel de la lógica tanto en aplicaciones a desarrollos concretos de IA (por ejemplo, en los desarrollos sobre representación del conocimiento) como fuente de motivación para nuevos resultados lógicos (por ejemplo, la lógica no monótona). Tanto los resultados clásicos como diversos sistemas no clásicos son usados de manera sistemática. La lógica computacional y el razonamiento automático son muestras claras en este sentido. En contraste, en la IA Conexionista las vinculaciones con la lógica clásica parecen ser menos reconocidas y sólo se suele

---

2 Véase Balzer, W., Moulines, C. U. y Sneed, J. D., *An Architectonic for Science. The Structuralist Program*, Dordrecht-Boston, D. Reidel, 1987.

3 El artículo original es Rosenblatt, F., "The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain" en *Psychological Review*, vol. 65, No. 6, 1958, pp. 386-408 y hay muchas exposiciones. Entre ellas véase, por ejemplo, Haykin, S., *Neural Networks. A Comprehensive Foundation*, New York, Macmillan College Publishing Company, 1992 y Wasserman, P., *Neural Computing. Theory and Practice*, New York, Van Nostrand Reinhold, 1989.

señalar la relación con la lógica borrosa (*fuzzy*). Sin embargo, hay algunos problemas y desarrollos concretos en la IA Conexionista donde la vinculación con la lógica clásica es explícita y fructífera, tal como veremos seguidamente. La *caracterización de los modelos* de las diversas teorías sobre redes neurales contribuye a identificar con claridad cada teoría, a establecer relaciones entre ellas y a facilitar su clasificación. También sirve como punto de partida para examinar algunos problemas metodológicos importantes.

Comencemos por caracterizar las redes más fundamentales: los perceptrones[4] de una y dos capas. Los perceptrones son redes basadas en elementos neurales idealizados y sencillos denominados "nodos de McCulloch–Pitts" o "nodos MCP". Estos elementos contienen algunos de los componentes fundamentales de las redes neurales más complejas y han jugado un importante papel en la construcción de los primeros computadores, específicamente en el EDVAC por parte de von Neumann[5]. McCulloch y Pitts desarrollan un cálculo lógico que exponen en su famoso artículo de 1943[6] y que constituye el punto de partida de la IA Conexionista. Entre las hipótesis fundamentales de este cálculo, el cual ha tenido un especial interés para los lógicos, se encuentran las siguientes: a) la presentación de la actividad neuronal se puede representar de manera binaria, b) la función de activación (la función que determina de qué modo se activa o dispara (*fires*) la red) es una función de umbral[7] y c) la teoría proporciona definiciones e interpretaciones de las funciones de verdad o funciones booleanas de la lógica clásica $\neg$, $\wedge$, $\vee$ basándose en la función de activación. Se ha demostrado que el cálculo resultante es un modelo del álgebra booleana.

---

4 El término es de Rosenblatt; véase Rosenblatt, op. cit.

5 Véase Haykin, *op. cit.*, p. 36.

6 McCulloch, W. y Pitts, W.,"A Logical Calculus of the Ideas Immanent in Nervous Activity" reimpreso en M. Boden (Ed.), *The Philosophy of Artificial Intelligence*, Oxford University Press, 1990, pp. 22-39.

7 Vale 1 o 0 dependiendo si la unidad aritmética de la red devuelve un valor mayor o menor a un cierto valor umbral. Es una función continua pero no es diferenciable (véase el axioma (11) de la p. 115).

Los nodos MCP son "elementos neurales idealizados" pues comparten sólo *algunos* aspectos de las neuronas biológicas descartando otros. El perceptrón de una capa (de procesamiento) es básicamente un nodo MCP en el cual se realizan ciertos procesamientos de información adicionales. Concatenando o uniendo perceptrones de una capa de una manera específica (que explicaremos posteriormente) podemos obtener perceptrones de varias capas de procesamiento, los cuales tienen mayor poder computacional y mayor rango de aplicaciones. Ahora bien, como afirma Wasserman,

"Despite the limitations of perceptrons, they have been extensively studied (if not widely used). Their theory is the foundation for many other forms of artificial neural networks and they demonstrate important principles. For these reasons, *they are a logical starting point for a study of artificial neural networks*"[8].

En la próxima sección definimos el *marco conceptual* de la teoría, a saber, el conjunto de todas las posibles entidades que pueden satisfacer los postulados de la teoría o conjunto de *modelos potenciales* de la misma[9].

---

8 Wasserman *op. cit.*, p. 29, subrayado mío (L.Z.M.).

9 Siguiendo el "*language free approach*" de Suppes continuado por la concepción estructuralista, no indicamos explícitamente el lenguaje formal empleado al definir los modelos de la teoría. Este modo de proceder parece ser más similar al álgebra universal que a la teoría de modelos propiamente dicha. De allí que Stegmüller y otros autores hablan de una teoría *informal* de modelos. No obstante, Rantala y Pearce en los años 80' formularon una interesante propuesta: entender esta caracterización de los modelos al margen de un lenguaje formal específico como parte de una *generalización* de la noción de modelo la cual es aplicable a *varios* clases de lenguajes formales simultáneamente. Este punto importante requiere un tratamiento detallado que proporcionaremos en un trabajo futuro. Por ahora conviene fijar la razón de esta estrategia: al no indicar de manera explícita el lenguaje formal subyacente podemos dirigir los esfuerzos más directamente a los axiomas *específicos* de la teoría, esto es, a axiomatizar *directamente* la teoría *presuponiendo* la lógica y la matemática necesaria (en este caso, el álgebra lineal de los espacios finito–dimensionales). La justificación de este proceder es pragmática: una vez establecidos los axiomas propios usando teoría *intuitiva* de conjuntos (en una primera etapa) podemos luego reformular la reconstrucción sobre la base de un lenguaje *formal* de teoría de conjuntos de primer orden (en una segunda etapa).

## *§ 2 Marco conceptual: modelo potenciales*

El marco conceptual de la teoría básica está representado por el conjunto $M_P$ de modelos potenciales, el cual contiene las tipificaciones y caracterizaciones matemáticas de los términos primitivos de la teoría. Definimos este conjunto mediante el predicado conjuntista $x \in M_P(Percep_1)$ o "*x* es un ***modelo potencial de un perceptrón de una sola capa***", de la siguiente manera[10]:

*DEF. 1:* $x \in M_P(Percep_1)$ o "*x* es un ***modelo potencial de un perceptrón de una sola capa***" $\Leftrightarrow_{Df}$ existen $\Gamma$, $E$, $j$, $C$, $\boldsymbol{x}$, $\boldsymbol{a}$, $\boldsymbol{w}$, $U$, $T$, $\varphi$, $y_D$, $y_A$ tales que

(1) $x = \langle \Gamma, E, j, C, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{w}, U, T, \varphi, y_D, y_A \rangle$

(2) $\Gamma$ es un espacio muestral de un cierto espacio de probabilidad, con una distribución de probabilidad no definida.

(3) $E$ es un conjunto con exactamente n elementos: $E = \{e_1, ..., e_n\}$, $n \in \mathbb{N}$ y $n \geq 2$.

(4) $j$ es un índice constante en $\mathbb{N}$.

(5) $C$ es un conjunto finito y no vacío.

(6) $\boldsymbol{x}$ es una función vectorial $\boldsymbol{x}$: $\Gamma \longrightarrow R^n$ ($n \in \mathbb{N}$ y $n \geq 2$) y $\boldsymbol{x} \in C$.

(7) $\boldsymbol{a}$ es una función binaria $\boldsymbol{a}$: $X \longrightarrow I \times I$, donde $I = \{0, 1\}$ y $X = \{x_i\}$ es el conjunto de componentes del vector $\boldsymbol{x}$, y $\boldsymbol{a} \in C$.

(8) $\boldsymbol{w}$ es una función vectorial $\boldsymbol{w}$: $\{a_i\} \times \{j\} \times E \longrightarrow R^n$ ($n \in \mathbb{N}$ y $n \geq 2$) donde $\{a_i\}$ es el conjunto de componentes del vector $\boldsymbol{a}$ y $\boldsymbol{w} \in C$.

(9) $U$ es una función $U$: $R^2 \longrightarrow R$ y $U \in C$.

(10) $T$ es una constante real no nula y $T \in C$.

---

10 'ℕ' y 'R' denotan, respectivamente, al conjunto de los números naturales y reales. Si f es una función con dominio D, la notación usual '$f \in C^n(D)$' indica que f es de clase $C^n$ en todo su dominio. Si n = 0, f es continua pero no derivable y si n > 0, f admite derivada de orden n.

(11) $\varphi$ es una función $\varphi$: $\mathrm{R} \longrightarrow \{0, 1\}$, $\varphi \in C^0(\mathrm{R})$ y $\varphi \in C$.
(12) $y_D$ es una función $y_D$: $\Gamma \longrightarrow \mathrm{R}$
(13) $y_A$ es una función $y_A$: $\{0, 1\} \times E \longrightarrow \mathrm{R}$ y $y_A \in C$.

INTERPRETACIÓN BÁSICA:

- $\Gamma$ representa el ***ambiente*** en el que está inmersa la red (y el cual provee al perceptrón de las señales de entrada; se trata de un conjunto estocástico).
- $E$ representa un conjunto discreto de ***estados*** que el sistema toma durante el proceso de funcionamiento de la red.
- $j$ representa la ***unidad de procesamiento o neurona artificial*** de la red.
- $C$ representa la ***capa de procesamiento*** de la red.
- $x$ representa el ***vector de entrada*** cuyas componentes a su vez representan *las* ***señales de entrada*** a la neurona artificial (también interpretado como el conjunto de ***estímulos*** para la red).
- $a$ representa el vector cuyas componentes a su vez representan las ***unidades asociativas o cajas lógicas*** del sistema.
- $w$ representa el ***vector de peso sináptico***. Cada una de sus componentes representa una *medida de la intensidad de la conexión entre las unidades asociativas y la neurona artificial j*.
- $U$ representa la unidad sumadora o ***unidad aritmética*** del sistema.
- $T$ representa el valor ***umbral***.
- $\varphi$ representa la ***función de activación*** de la red.
- $y_D$ representa la señal de ***salida deseada*** u ***objetivo*** (*target*); sus componentes representan las ***señales de salida*** de la neurona artificial.
- $y_A$ representa la señal de ***salida actual***, es decir, aquella obtenida por los cálculos realizados mediante la unidad aritmética y la función de activación. (También interpretada como el conjunto de ***respuesta*** de la red).

OBSERVACIÓN 1 (sobre los modelos potenciales):

a) El concepto de aprendizaje (que forma parte de las *especializaciones* del núcleo fundamental)[11] juega un papel fundamental en las redes neurales; de hecho, a este aspecto se debe buena parte del interés que éstas han despertado. Ahora bien, el tipo de aprendizaje que se usa en los perceptrones es el aprendizaje *supervisado*, y uno de sus más importantes características es la suposición del carácter *estocástico* de ese proceso. De acuerdo a la descripción matemática del aprendizaje supervisado hecha por Vapnik, el ambiente o entorno $\Gamma$ provee a la red de un vector de entrada $x$ mediante una distribución de probabilidad p($x$) fija pero desconocida. Por tanto, los métodos de la estadística no paramétrica son requeridos aquí. Nótese que en el axioma (2) la cardinalidad de $\Gamma$ no está restringida, por tanto, ésta puede ser un espacio de probabilidad discreto o continuo.

b) Respecto al concepto de estado es conveniente aclarar lo siguiente: dados dos estados $e_i$ y $e_{i+1}$, lo que es relevante para la teoría es registrar y comparar los conjuntos de valores en ambos estados y *no* lo que ocurre *durante la transición* de un estado a otro. Al conjunto $E$ de estados se aplican las mismas consideraciones que a las máquinas de Turing, con las cuales pueden simularse el funcionamiento del perceptrón.

c) Las unidades asociativas o cajas lógicas son usadas para formalizar las ideas de Rosenblatt y los perceptrones máscara (*mask perceptrons*) estudiados por Minsky y Papert[12]. Estas unidades permiten realizar cierto *preprocesamiento* de las entradas.

d) No todas las funciones modifican sus valores al pasar de un estado a otro; por ejemplo, los pesos sinápticos modifican sus valores pero el vector de entrada no. Es en este

---

11 Véase *infra* p. 9.

12 Cfr. Minsky, M. y Papert, S., *Perceptrons. An Introduction to Computational Geometry*, Cambridge (Massachusetts), Expanded Edition, The MIT Press, 1988.

sentido que se dice frecuentemente que la función $\boldsymbol{w}$ es una función *adaptativa*.

## *§ 3 Leyes o axiomas fundamentales: modelo actuales.*

El conjunto de modelos actuales M se obtiene a partir del conjunto de modelos potenciales $M_P$ postulando las *leyes o axiomas propios o fundamentales* de la teoría. En este caso tenemos una ley muy general la cual es satisfecha en todas las redes neurales[13] y que denominamos "ley fundamental", y otras dos leyes que establecen las diferencias claves entre el perceptrón y las otras redes neurales. La ley fundamental establece la relación funcional que existe entre la salida actual $y_A$ por un lado y la función de activación $\varphi$ y cierto subconjunto $E_0$ de $E$ por otro. En la forma tradicional ésta se establece como $y_A = y_A(\varphi, E_0)$ o como $y_A = f^{y}_{A}(\varphi, E_0)$ en la forma lógicamente más adecuada que es usual en la metodología estructuralista. Según esta última versión, la función $y_A$ depende de un *funcional* $f^{y}_{A}$ que tiene como argumentos a la función $\varphi$ y al conjunto $E_0$. *Sin este postulado la red sería totalmente inútil debido a que no habría manera de relacionar la salida de la red con sus operaciones internas* (determinadas por la función $\varphi$).

Las otras dos leyes establecen la forma que toma la unidad aritmética $U$ y la función de activación $\varphi$. $U$ se define como el producto interno $\boldsymbol{a.x}$ y $\varphi$ como una función de umbral. En otras redes neurales ambas funciones se definen de forma diferente. Por ejemplo, la unidad aritmética $U$ toma una forma distinta en el Cognitrón y la función de activación se define en Retropropagación como una función sigmoide $\varphi(U) = 1/1+e^{-aU}$ (donde el parámetro a representa la pendiente de la función) la cual es diferenciable. Con estas consideraciones en mente podemos definir los modelos actuales del

---

13 Más en concreto, esta ley es válida para una teoría general de perceptrones multicapa (que incluye teorías más específicas tales como la Retropropagación).

perceptrón así:

DEF. 2: $x \in M(Percep_1)$ o "$x$ es un ***(modelo actual de un) perceptrón de una sola capa***" $\Leftrightarrow_{Df}$ existen $\Gamma$, $E$, $j$, $C$, $\mathbf{x}$, $\mathbf{a}$, $\mathbf{w}$, $U$, $T$, $\varphi$, $y_D$, $y_A$ tales que

(1) $x = \langle \Gamma, E, j, C, \mathbf{x}, \mathbf{a}, \mathbf{w}, U, T, \varphi, y_D, y_A \rangle \in M_P(Percep_1)$

(2) $U = \sum_{i=1}^{n} w_i a_i$

(3) $\varphi(U) = \begin{cases} 1 \text{ si } U \geq 0 \\ 0 \text{ si } U < 0, \text{ y la superficie } U = T \text{ existe, i.e.} \\ \left\{ (w, a) / \sum_{i=1}^{n} w_i a_i = T \right\} \neq \varnothing. \end{cases}$

(4) $\exists E_O \subseteq E (y_A = f^{y}_{A}(\varphi, E_O) \wedge \exists E_m \subseteq E_O (n \geq m \wedge y_A(e_m) = y_D))$.

OBSERVACIÓN 2 (sobre los modelos actuales):

a) En el axioma (3) se establece una de las restricciones más importantes y fuertes de la teoría: el perceptrón de una capa sólo puede clasificar entre conjuntos ***linealmente separables***. Esto es, la red sólo puede clasificar conjuntos de entrada que son separables por una recta (en R), por un plano (en $R^2$) o por un hiperplano (en $R^n$, $n \geq 3$) el cual se denomina *superficie de decisión*. La red no puede clasificar conjuntos conexos o conjuntos separables de un modo no lineal. Podemos definir una *separación o desconexión* del conjunto $X = \{x_1, ..., x_n\}$ de entradas (componentes del vector de entrada) mediante un par de subconjuntos $A$, $B$ de $X$ tal que $A \cup B = X$, $A \cap B = \varnothing$ y donde tanto $A$ como $B$ son ambos abiertos (o ambos cerrados) en $X$, y podemos pasar ahora a definir una separación o desconexión entre $A$ y $B$ como *lineal* cuando puede ser establecida mediante una recta, plano o hiper-

plano dado por la ecuación $U = T$. Si esta ecuación genera una superficie de decisión, estamos ante un modelo actual de la red, si el modelo potencial genera una contradicción, entonces no existe la superficie de decisión lo cual indica que ese modelo *potencial* no es ***expandible*** a un modelo *actual* de la misma.

b) En el elemento teórico básico sólo postulamos que la igualdad $y_A = y_D$ se obtiene en un número finito de pasos pero no proporcionamos ningún algoritmo concreto para lograrlo. Algoritmos de este tipo se postulan en las ***especializaciones*** del elemento teórico básico, en forma de *algoritmos de aprendizaje* proporcionados por la teoría[14].

c) Podemos usar distintas técnicas de teoría de grafos para representar la arquitectura y modo de funcionamiento de una red, por ejemplo mediante diagramas de bloque, grafos dirigidos y otros. Pero no podemos *identificar*, en la reconstrucción lógica expuesta, una red neural con un grafo de este tipo; *consideramos como modelo de la teoría a un conjunto de estados que satisfacen los axiomas*; la arquitectura por sí sola no caracteriza suficientemente a un modelo de la teoría. Si el grafo en cuestión incluye la descripción de los estados que toma la red durante su funcionamiento, entonces sí se trata de un modelo de la teoría.

La arquitectura general de la red se ilustra en el gráfico

---

[14] Así como en mecánica clásica la *ley fundamental* es la segunda ley de Newton y leyes más específicas como la de Hooke forman parte de las *especializaciones* de este núcleo básico, en nuestro caso el axioma DEF. 2-(4) es la ley general y la *regla delta* aparece en una especialización de este núcleo básico. Otro ejemplo usual en termodinámica es tomar una ecuación general de estado como ley general y considerar las leyes de los gases ideales o de van der Waals como parte de la especialización. La demostración del teorema de convergencia así como la definición de las estructuras conjuntistas correspondientes a los distintos *algoritmos de aprendizaje* aparecen expuestos con todo detalle en Zerpa, L., *Una aproximación lógica a la Inteligencia Artificial Conexionista,* libro de próxima publicación por la Comisión de Estudios de Postgrado, Universidad Central de Venezuela, Caracas.

siguiente (véase Fig. 1):

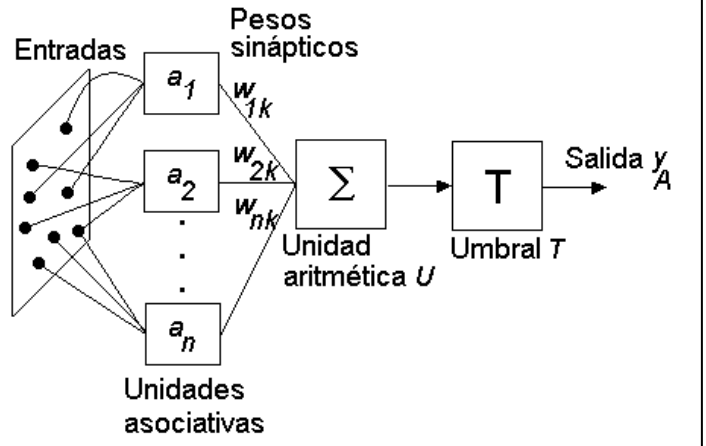


Fig. 1 Perceptrón de una capa. Caso general.

RESUMEN DE LA INTERPRETACIÓN GEOMÉTRICA (EN $R^n$) DE LOS MODELOS ACTUALES (CASO GENERAL):

*En el caso general un modelo del perceptrón de una capa es un conjunto finito de estados en los cuales se lleva a cabo una clasificación de los componentes del vector de entrada* ***x***. La superficie de decisión $U = T$ divide el conjunto de componentes de ***x*** en dos conjuntos $X_1 = \{x/\varphi(U) = 0\}$ y $X_2 = \{x/\varphi(U) = 1\}$ los cuales son linealmente separables. El vector ***x*** puede ser binario o continuo.

DEF. 3: Sea un $x \in M(Percep_1)$. El perceptrón $x$ ***se activa*** si $y_A = 1$ y ***se inhibe*** si $y_A = 0$ en $x$.

DEF. 4: Un perceptrón de una capa $y \in M(Percep_1)$ es ***directo*** si $\boldsymbol{a} = \boldsymbol{x}$ en y[15].

15 Si los valores de las unidades asociativas coinciden con las entradas se presenta este caso particular tan común.

DEF. 5: Un perceptrón de una capa es binario si y sólo si $x \in M(Percep_1) \wedge Rec(x) = Rec(y_D) = Rec(y_A) = \{0, 1\}$.

RESUMEN DE LA INTERPRETACIÓN GEOMÉTRICA (EN $R^n$) DE LOS MODELOS ACTUALES (CASO PARTICULAR: PERCEPTRÓN BINARIO):

Sea un modelo $x \in M(Percep_1)$ de un perceptrón binario. Por el axioma (3) de la DEF. 2 tenemos el siguiente par de desigualdades

$$\varphi(U) = 1 \Leftrightarrow w_1x_1 + ... + w_nx_n \geq T$$
$$\varphi(U) = 0 \Leftrightarrow w_1x_1 + ... + w_nx_n < T.$$

Consideremos el caso $w_1x_1 + ... + w_nx_n = T$. Esta ecuación puede ser interpretada geométricamente como un *hiperplano $\pi$ que divide los valores de $\boldsymbol{x}$ para los cuales $\varphi(U) = 1$ de aquellos valores de $\boldsymbol{x}$ para los cuales $\varphi(U) = 0$ en el hipercubo unidad*.

EJEMPLO (INCOMPATIBILIDAD (*NAND* o BARRA DE SHEFFER) $|$): Sea un $a_0$ tal que $a_0 = <\Gamma, E, j, C, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{w}, U, T, \varphi, y_D, y_A>$, T = –0,5; los pesos están en [–1, 1] y $\boldsymbol{x}$ y $y_D$ son dados por la tabla de verdad de la función booleana *NAND* la cual denotamos por $f^|$ (donde '$|$' es la barra de Sheffer):

$$f^|(1, 1) = 0,$$
$$f^|(1, 0) = 1,$$
$$f^|(0, 1) = 1,$$
$$f^|(0, 0) = 1.$$

A partir de los postulados de la teoría y los valores dados podemos obtener lo siguiente:

$E = \{e_1, e_2\}$ (donde cada $e_i$ puede interpretarse como un paso de computación), $j = 1$ (hay sólo una unidad de procesa-

miento), $\boldsymbol{a} = \boldsymbol{x} = (x_1, x_2)$ (el perceptrón es directo[16]) y las entradas son $x_1 = (1, 1, 0, 0)$, $x_2 = (1, 0, 1, 0)$; $\boldsymbol{w} = (w_1, w_2)$ y $\varphi(U) = 0 \Leftrightarrow w_1x_1 + w_2x_2 < -0{,}5$. Por sustitución obtenemos:

(1) $w_1 + w_2 < -0{,}5$
(2) $w_1 \geq -0{,}5$
(3) $w_1 \geq -0{,}5$
(4) $0 \geq -0{,}5$.

A partir de (2), (3), (4) y la restricción $w_1, w_2 \in [-1, 1]$, podemos obtener infinitas soluciones, por ejemplo, $w_1 = -0{,}3$ y $w_2 = -0{,}4$. En base a estos valores la superficie de decisión $U = T$ existe y es la recta $(-0{,}3)x_1 + (-0{,}4)x_2 = -0{,}5$ la cual separa los puntos con valor 1 de aquellos con valor 0 en el cuadrado unidad[17]. Por tanto $a_0$ es un modelo de la red, esto es, $a_0 \in \mathrm{M}(\mathrm{Percep}_1)$. Si tomamos la misma función en 3 variables obtenemos como superficie de decisión a un plano que separa puntos en el cubo unidad (1, 1, 1), ..., (0, 0, 0). Lo mismo vale para $n > 3$ variables.

CONTRAEJEMPLO (DISYUNCIÓN EXCLUSIVA (*XOR*)):

Sea un $b_0 = \langle \Gamma, E, j, C, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{w}, U, T, \varphi, y_D, y_A \rangle$ con los mismos valores de $T$ y $\boldsymbol{x}$, y $y_D$ es dada por la tabla de verdad de la función booleana disyunción exclusiva (exclusive OR, XOR) $f^{\vee}$:

$f^{\vee}(1, 1) = 0$,
$f^{\vee}(1, 0) = 1$,
$f^{\vee}(0, 1) = 1$ y
$f^{\vee}(0, 0) = 0$.

A partir de los postulados de la teoría y los valores dados podemos obtener lo siguiente:

16 Véase la DEF. 4.
17 Es decir el cuadrado definido por los puntos (1, 1), (1, 0), (0, 1) y (0, 0).

$\varphi(U) = 1 \Leftrightarrow w_1x_1 + ... + w_2x_2 \geq 0{,}5$ y $\varphi(U) = 0 \Leftrightarrow w_1x_1 + ... + w_2x_2 < 0{,}5$.

Por sustitución obtenemos:

(1) $w_1 + w_2 < 0{,}5$
(2) $w_1 \geq 0{,}5$
(3) $w_1 \geq 0{,}5$
(4) $0 < 0{,}5$

Aplicando la propiedad $(a > c \wedge b > c) \Rightarrow a + b > c$, obtenemos $w_1 > T$ y $w_2 > T$, lo cual implica que $w_1 + w_2 > T$ en contradicción con (1). Como ***no*** existe la superficie de decisión $U = T$ entonces $\{(w, x)/\Sigma w_i x_i = T\} = \varnothing$, por tanto $b_0$ ***no*** es modelo de $\text{Percep}_1$, esto es, $b_0 \not\in \text{M}(\text{Percep}_1)$. El mismo resultado se obtiene otra función booleana: el bicondicional o equivalencia material $\leftrightarrow$. En consecuencia, puede demostrarse que de 16 funciones booleanas definibles en 2 variables el perceptrón de una capa puede "representar" (véase la DEF. 6, p. 14) 14 funciones booleanas linealmente separables y falla en la representación de las únicas 2 que no lo son.

OBSERVACIÓN 3 (Sobre el problema de la disyunción exclusiva (*XOR problem*)):

1) Como acabamos de ver, el perceptrón de una capa no es capaz de representar funciones no linealmente separables como la disyunción exclusiva y el bicondicional. A esto se le ha llamado en la literatura, después de Minsky y Papert, el "*problema de la disyunción exclusiva*" ("*XOR problem*"). Buena parte de la evolución de las redes neurales ha sido motivada por la necesidad de superar esta gran limitación[18].

[18] En efecto, al considerar funciones booleanas con más de 2 variables el número de funciones no linealmente separables crece muy rápidamente. Por ejemplo, para el caso n = 3 menos de la mitad de las funciones son linealmente separables (128 contra 104) y la proporción es muchísimo menor para n = 5:

2) A primera vista luce muy extraño el hecho que la red pueda representar la incompatibilidad $|$ mientras que es incapaz de representar la disyunción exclusiva, siendo $\{|\}$ un conjunto adecuado de conectivas[19].
3) A continuación definimos, en analogía con las máquinas de Turing, la representación de valores de una función mediante un perceptrón binario.

DEF. 6: Un ***perceptrón binario*** puede ***representar los valores de una función binaria f*** $\Leftrightarrow_{Df}$ existe un modelo $a_O \in M(Percep_1)$ tal que si tomamos como vector de entrada el dominio de la función, i.e. $x$ = Dom($f$), entonces podemos obtener la salida de la red como su recorrido o rango: $\exists e_m \in E[y_A = Rec(f)]$.

## *§ 4 Solución al problema de la representación de funciones booleanas no linealmente separables: concatenación de perceptrones de una capa mediante compuertas lógicas*

Siguiendo a Aleksander and Morton[20] podemos demostrar que si bien un perceptrón de una capa no puede representar funciones linealmente no separables, uno de ***dos capas*** sí puede representar *cualquier* función de este tipo. ¿Cómo? *Concatenando o uniendo mediante compuertas lógicas a dos perceptrones de una capa*. En efecto, sean las funciones booleanas disyunción inclusiva $f^{\vee}$ e incompatibilidad $f^{|}$:

$$f^{\vee}(1, 1) = 1,$$
$$f^{\vee}(1, 0) = 1,$$
$$f^{\vee}(0, 1) = 1,$$

---

de un total de 4.300.000.000 funciones hay 2.149.905.428 no linealmente separables contra 94.572 que sí lo son. Cfr. Wasserman, op. cit., p. 34.

19 Es decir, que toda función booleana puede expresarse mediante una forma enunciativa en la que sólo aparece $|$ y las variables enunciativas. Este problema se analiza con detenimiento en Zerpa op. cit., cap. 3, secc. 6.

20 Cfr. Aleksander, I. y Morton, H., *An Introduction to Neural Computing*, Londres, Chapman and Hall, 1990, cap. 3, secs. 3.4 y 3.5.

$f^{\vee}(0, 0) = 0$, y
$f^{|}(1, 1) = 0$,
$f^{|}(1, 0) = 1$,
$f^{|}(0, 1) = 1$,
$f^{|}(0, 0) = 1$.

Si tomamos la conjunción de ambas funciones $f^{\vee}$ e incompatibilidad $f^{|}$ obtenemos la deseada representación de la disyunción exclusiva:

$(f^{\vee} \wedge f^{|})(1, 1) = 0$,
$(f^{\vee} \wedge f^{|})(1, 0) = 1$,
$(f^{\vee} \wedge f^{|})(0, 1) = 1$,
$(f^{\vee} \wedge f^{|})(0, 0) = 0$.

Esto es, mediante la conjunción $f^{\vee} \wedge f^{|}$ de las funciones $f^{\vee}$ y $f^{|}$ obtenemos una función booleana que es equivalente a la disyunción exclusiva. En términos de la función de activación consideremos a $f^{\vee}$ definida por $x_1 + x_2 > 0{,}5$ de lo cual obtenemos $2 > 0{,}5$; $1 > 0{,}5$; $1 > 0{,}5$ y $0 < 0{,}5$. Igualmente, si consideramos a $f^{|}$ definida por $x_1 + x_2 > 0{,}5$ podemos obtener $-2 < -1{,}5$; $-1 < -1{,}5$; $-1 < -1{,}5$ y $0 > -1{,}5$.

Más aún, mediante técnicas usuales en lógica digital podemos generalizar este resultado. Específicamente, la *primera y segunda forma normal* de un *circuito lógico* le permite al perceptrón de una capa representar ***cualquier*** función booleana en cualquier número de variables. La arquitectura de la red resultante se puede describir brevemente así: si la red tiene 2 entradas ($\boldsymbol{x}$ es k-dimensional) las capas se distribuyen del siguiente modo: las $k$ entradas ingresan a la primera capa que tiene $2^{(k-1)}$ nodos MCP y las correspondientes $2^{(k-1)}$ salidas entran a la segunda capa que consta de un solo nodo MCP. Este último nodo computa los resultados obtenidos en los nodos de la primera capa. En la Fig. 2 se ilustra el caso k=2. Nótese que desde el punto de vista metodológico, las redes

neurales tienen una diferencia notable respecto a otras teorías. Cuando concatenamos redes simples (de una capa) para obtener redes más complejas (de dos o más capas) el resultado es una red más poderosa computacionalmente hablando pues ella puede resolver problemas que son insolubles para cada red simple tomada por separado. En contraste, cuando concatenamos sistemas físicos de otro tipo, por ejemplo sistemas termodinámicos, eso no ocurre: el sistema concatenado parece resolver el mismo tipo de problemas que pueden resolverse con cualquiera de los sistemas simples tomados por separado[21].

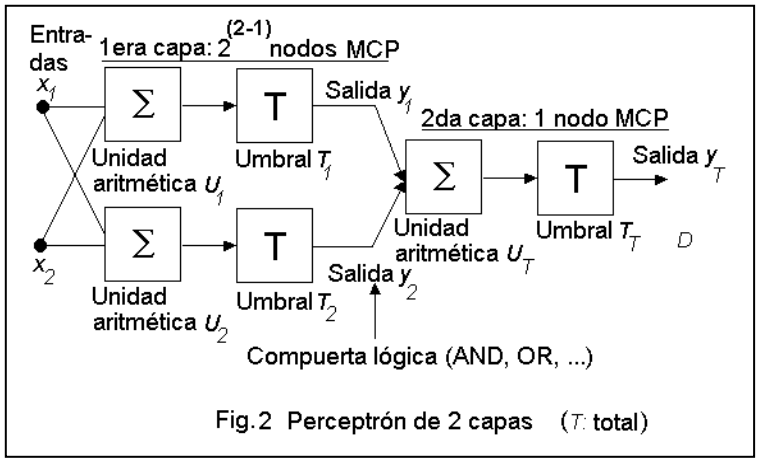


Fig.2 Perceptrón de 2 capas (*T*: total)

Al ***concatenar*** redes de una capa para producir redes de dos capas hemos dado un paso fundamental desde el punto de vista *lógico y metodológico*: hemos empezado a establecer ***relaciones intermodélicas***, es decir, relaciones entre los distintos modelos. Y como es sabido, estas relaciones se representan en la metodología estructuralista mediante el importante concepto de ***condición de ligadura*** introducido por Sneed a co-

21 Véase Zerpa, L., "El rol de las condiciones de ligadura en la ciencia y en la filosofía de la ciencia recientes" ponencia presentada en el V Congreso Nacional de Filosofía, Caracas, noviembre de 1999 (mimeografiado).

mienzos de los años 70'[22]. Este concepto se ha usado para representar las relaciones de *solapamiento* entre las distintas aplicaciones de una teoría empírica. Tres ejemplos notables de estos solapamientos son las "ligaduras de igualdad", las "ligaduras de constantes" y las ligaduras de concatenación". Ejemplo concreto de las primeras: en mecánica clásica la masa de una partícula no varía con la velocidad, luego, si una partícula *p* aparece en varios modelos su valor será el mismo en todos ellos. Ejemplo de las segundas: la invarianza de cualquier constante física (como la constante de los gases ideales o la constante de gravitación). En las redes neurales estos dos primeros tipos de condiciones de ligadura no parecen tener una gran importancia pues las funciones y las constantes varían de modelo en modelo (es decir, de una aplicación a otra)[23]. En cambio, las ligaduras de concatenación *sí* tienen un lugar muy destacado pues mediante ellas podemos representar la concatenación de perceptrones mediante compuertas lógicas y representar funciones no linealmente separables como acabamos de ver.

Ahora bien ¿de qué manera podemos representar la concatenación de perceptrones como ligaduras de concatenación? Según la interpretación propuesta, en cada modelo potencial $x \in M_P(Percep_1)$ se caracteriza una unidad aritmética *U*, y cada red neural de una capa se puede describir mediante un modelo actual *y* de la teoría, $y \in M(Percep_1)$. Cada modelo actual **como un todo** se hace corresponder con un perceptrón de una capa. La pregunta clave aquí es ésta: ¿cómo podemos unir o concatenar a los sistemas descritos por los modelos *x* y *x'*? Como acabamos de ver, podemos unirlos mediante una *compuerta lógica* ($\wedge$ *(AND)*, $\vee$ *(OR)*, | *(NAND)*, etc.) de tal modo que la unidad aritmética del mo-

---

22 Véase Balzer *et al*., op. cit.

23 Esto se cumple tanto para los modelos tal como los hemos definido como para las especializaciones del núcleo básico que se obtienen introduciendo los algoritmos de aprendizaje (de incremento fijo y regla delta). Por ejemplo, la constante de aprendizaje varía de aplicación en aplicación. Cfr. Zerpa, *Una aproximación lógica...*, op. cit., cap. 3.

delo $x$ se conecta a la unidad aritmética del modelo $x'$ mediante esa compuerta lógica (véase la Fig. 2). De este modo la salida $y_1$ de la primera red (modelo $x$) y la salida $y_2$ (modelo $x$) de la segunda red forman las entradas de la red compuesta $xox'$ (la notación es de Balzer *et al*, op. cit.[24]) de tal modo que la unidad aritmética $U_T$ calcula los resultados obtenidos en $x$ y $x'$. En analogía con la concatenación de sistemas termodinámicos, la operación de concatenación de estados la denotamos por '$o$' y en consecuencia denotamos por $zoz'$ o por $z''$ el sistema compuesto por la concatenación de los estados $z$ y $z'$ (donde Z, $Z'$ y $Z''$ son conjuntos de estados tales que $z \in Z$, $z' \in Z'$, $z'' \in Z''$ y $o: Z \times Z' \rightarrow Z''$). Ya que los modelos son conjuntos de estados que cumplen ciertas condiciones, entonces podemos concatenar modelos mediante la operación $o$ de tal modo que podemos escribir $x \; o \; x' = x''$ para representar el modelo concatenado, como ya mostramos en la Fig. 2, donde $x$ y $x'$ están en la primera capa y $x''$ en la segunda capa.

Ahora bien, si $x, x' \in M_P(Percep_1)$ y $x$ y $x'$ están en los estados $s$ y $s'$ respectivamente, y $x$ tiene asignada una unidad aritmética $U(s)$ y $x'$ tiene asignada una unidad aritmética $U'(s')$[25], entonces, $sos'$ tiene asignada una unidad aritmética $U''(sos')$. Ahora bien, ¿qué *forma* tiene $U''(sos')$? $U''(sos')$ tiene la forma general $U''(sos') = f(U(s), U'(s'))$ donde $f$ es una función booleana o veritativa. Por tanto, la condición de ligadura puede definirse así:

(**) $\forall x, x'x'' \in X \subseteq M_P\,(Percep_1)\; [\; (x'' = xox' \wedge x, x', x''$ *están en los estados s, s' y s'' respectivamente)* $\rightarrow U''\,(sos') = f(U(s), U'(s')) \wedge f$ *es una función booleana]*

Otras estructuras conjuntistas importantes que podemos encontrar son los vínculos interteóricos, especialmente los vínculos con las neurociencias (tópico que merece un desarrollo detallado). Si consideramos que el núcleo estructural

---

24 La tilde "" sólo indica que se trata de otro modelo, no se trata de una derivada.

25 Véase la nota anterior sobre la notación $U'(s')$; la tilde en la función $U$ indica ella se define en el modelo $x'$; no hay ninguna operación de derivación aquí.

está formado, básicamente, por los modelos, la ligadura de concatenación y los vínculos interteóricos, entonces entre las especializaciones de este núcleo podemos incluir los algoritmos de aprendizaje. De este modo, diversas aplicaciones relacionadas con el reconocimiento de patrones pueden obtenerse como modelos de la teoría de un modo bastante natural[26].

Instituto de Filosofía,
Universidad Central de Venezuela

[26] Estos desarrollos aparecen en Zerpa, op. cit., caps. 3 al 5.

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 127–134

NOTAS Y DISCUSIONES

JESÚS F. BACETA V.

# SOBRE FORMA LÓGICA, ESTRUCTURA PROFUNDA, ENUNCIADOS DE CREENCIA Y ONTOLOGÍA.
## Notas sobre un artículo de R. Bravo.

**Resumen:** Se señalan ciertas aclaratorias de carácter formal y algunas críticas a partir del desarrollo que propone R. Bravo en su artículo: "El compromiso ontológico de los lenguajes naturales".

**Palabras claves:** Intensión, lenguaje natural.

**Abstract:** In this article, are pointed out certain formal character explanations and some critics by taking as starting point the ideas that R. Bravo develops in his article: El compromiso ontológico de los lenguajes naturales".

**Keywords:** Intention, natural language.

Me limitaré aquí a señalar ciertos puntos de carácter formal y algunas aclaratorias a partir del desarrollo que propone R. Bravo en su artículo: *El compromiso ontológico de los lenguajes naturales*[1].

Bravo indica que: "los nombres propios permanecen como constantes irreductibles" (p. 52) lo cual es contrario a la razonable concepción de Russell según la cual los nombres propios pueden considerarse descripciones abreviadas, lo que equivale a decir que son analizables y, por tanto, reductibles; esto es, los nombres propios pueden ser sustituidos por descripciones definidas en las cuales no aparecen dichos nombres. En todo caso, para Russell sí habían nombres irreductibles y no eran precisamente los nombres propios, sino aquellos que sólo pueden aprenderse de manera ostensiva; esto es, cuando el nombre es el símbolo de algo experimentado. Este es el

1 Bravo, R.: *El compromiso ontológico de los lenguajes naturales,* en EPISTEME NS, N° 19, pp. 37 – 52.

caso de "verde", "rojo", etc. que Russell considera como nombres de cualidades simples y no universales.

Ahora bien, Quine propuso una paráfrasis mediante la cual las descripciones definidas son eliminables a favor de ciertas expresiones que contienen sólo variables[2], equiparables en su función a los pronombres relativos del español. Así tenemos un lenguaje sin términos singulares con el mismo poder expresivo que uno que contenga nombres, lo cual prueba que los nombres no constituyen un instrumento imprescindible para hablar acerca de objetos y que su completa eliminación no implica una pérdida del poder expresivo del lenguaje. De modo que, por ejemplo, el asunto russelliano de los "nombres aprendidos ostensivamente", se traduce al problema de

---

2 Su procedimiento fue a grandes rasgos el siguiente:
La 'designación' es referencia por medio de un término singular: «'Sócrates' designa a Sócrates»; la 'denotación' es referencia por medio de un término general, o predicado: «'conejo' denota a cada conejo». La designación se explica mediante el conocido expediente de equipararla con la denotación, esto es, eliminando los términos singulares a favor de los términos generales con la finalidad de llenar vacíos veritativos. Por ejemplo, en la oración 'Pegaso vuela', que carece de variables ligadas y que no es, en principio, ni verdadera ni falsa, se puede tratar el término 'Pegaso' como un término general, proporcionando la siguiente forma canónica que elimina la laguna veritativa de la oración original: '$(\exists x)(x$ = Pegaso y $x$ vuela)', la cual es falsa. En general, sea '$a$' un término singular y sea '$Fa$' la representación de una oración que contiene el término singular en cuestión. Quine propone parafrasear '$Fa$' como '$(\exists x)$ $(Fx \wedge a = x)$' y trata a la identidad con su parte izquierda '$\underline{a =} \; x$' como un nuevo término general o de predicado, digamos '$Gx$', que representa a un predicado que denota a un sólo individuo: «la cosa llamada 'Pegaso'», «el hombre llamado 'Sócrates'», auténticos términos generales. Aunque podría parecer una caracterización *ad hoc* el tratamiento de la designación como denotación, las ventajas son específicas: la mencionada intención de llenar lagunas veritativas, de por sí de una gran utilidad en el tratamiento de ciertos contextos donde los nombres no nombran nada, se logra trasladando el problema ontológico, que clásicamente recaía sobre los nombres, a las variables: «una teoría asume una entidad si y sólo si esta entidad debe incluirse entre los valores de las variables para que los enunciados afirmados en la teoría sean verdaderos» (Quine: "Acerca de lo que hay" [1948] en *Desde un punto de vista lógico,* cit., p. 154.) Se logra mayor economía en los análisis porque las leyes de la lógica se simplifican al no ser necesarias las eliminaciones de los cuantificadores y su respectiva ejemplificación por medio de términos distintos a las variables; con ello se eliminan, colateralmente, los supuestos de existencia que se aplican mediante la eliminación del cuantificador existencial o mediante la introducción del generalizador y se sugiere, además, una definición sintáctica de nombre: aquellas expresiones constantes que reemplazan a las variables y son reemplazadas por variables de acuerdo con las leyes lógicas usuales de la cuantificación.

cuáles son los estímulos que nos permiten asentir ante el proferimiento de unos sonidos que, desde el punto de vista formal, son representados por un predicado.

El discurso cotidiano no es, ciertamente, demasiado meticuloso en lo tocante a la ontología y, por consiguiente, es de esperar que una decisión basada en los pronombres relativos dé paso a un mundo excesivamente abigarrado. Las distintas paradojas que han surgido de la teoría de conjuntos nos dan la lección según la cual podemos evitar las contradicciones restringiendo los conjuntos que acepte un lenguaje; hay que restringir los universales que aceptamos en un lenguaje para evitar las paradojas.

Bravo insiste en que el criterio de Quine es normativo, por lo que funda su supuesto "criterio ontológico de los lenguajes naturales" en una interpretación descriptiva de la propuesta quineana, basada en cierta "relevancia del contenido intensional" que intenta distinguir entre 't' y 'cosa que es t', siendo 't' un término singular. Aquí las confusiones son considerables.

La paráfrasis del lógico, la llamada "forma lógica", y la paráfrasis del gramático, que algunos pueden llamar "estructura profunda", difieren no en calidad pero sí en detalles y propósitos. Las transformaciones de Quine son una austera y diáfana paráfrasis que no contiene términos singulares sino variables; la otra, la del lingüista, es algorítmicamente más eficiente y está llena de términos singulares complejos. Claro está, la paráfrasis lógica puede eliminar o no los términos singulares según cuál sea su propósito específico. Y, si dirigimos nuestro interés a la gramática española estamos condenados a descubrir, como se le reveló a Bravo, que la eliminación de términos singulares, con respecto a la gramática, carece de propósito. Considérese la distinción entre uso referencial y no referencial de los términos singulares. Muchos lingüistas consideran que tal distinción es vital para la apreciación de una lengua natural como el español, sin embargo, una paráfrasis lógica a lo Quine la hace totalmente imperceptible. Quine no formula su criterio ontológico para un despropósito y parece inadecuado confundir los propósitos de una paráfrasis endilgándole otro y llamar al primero "normativo" y, al segundo, "descriptivo". Tanto la forma lógica como la estructura profunda son paráfrasis de oraciones del lenguaje cotidiano, paráfrasis a las que recurrimos en virtud de ciertos propósitos de conveniencia técnica. Pero los propósitos no son los mismos: el lingüista o el gramático organiza la oración de una forma que pueda ser generada de la manera más eficiente por el árbol

gramatical; el propósito del lógico es organizar la oración en la forma que admita más eficientemente el cálculo lógico o muestre claramente sus implicaciones y afinidades conceptuales obviando falacias y paradojas. Bravo incurre en el error de pretender formar una nueva paráfrasis que sirva por igual a los diferentes propósitos del lingüista y del lógico. Esto no es así, aunque una paráfrasis pueda ser sugerente respecto de la otra.

Según Bravo, 't' es distinto de 'cosa que es t', siendo 't' un término singular, desde el punto de vista del contenido intensional. Está claro que desde el punto de vista extensional no existe diferencia alguna. Pero Bravo cree que desde el punto de vista intensional hay una diferencia sustancial y para sustentarlo dice:

> «La expresión formalizada más adecuada del enunciado del lenguaje ordinario 'algunos perros son blancos', sería, pues, algo así como:
>
> $(\exists p)$ $p$ es blanco,
>
> manteniendo la predicación sobre la variable definida del lenguaje natural 'perro', ya que la expresión original, contrariamente a Quine (y a Russell), no dice que "algunas cosas que son perros son blancas", sino, literalmente, que algunos *perros* son blancos; el enunciado del lenguaje natural no habla acerca de "cosas que son perros", sino acerca de perros, tipo específico de "cosa".» (p. 53).

Hay una especie de malentendido referente a la naturaleza de los cuantificadores en la propuesta de Bravo. Pareciera que cuantifica sobre una variable predicativa, con lo cual estaría en el terreno de una lógica de orden superior, pero Bravo afirma «…la formalización de 'algunos perros son blancos' como '$(\exists p)$ $p$ es blanco', cuantificando la variable definida '$p$', no comporta la reificación del término 'perro' como nombre de clase o atributo» (p. 54), esto es, según Bravo se encuentra en el campo de la lógica de primer orden. También pareciera que el operador de cuantificación tiene como alcance nombres y no variables, lo cual, a todas luces, es inadmisible. Pareciera indicar algo así como: "Existe perro, perro es blanco" tal como si dijéramos "Existe Juan, Juan es alto". Un cuantificador es inútil o vacío cuando su variable no se presenta; en pocas palabras, no se entiende sobre qué está cuantificando o realiza ciertas operaciones en una lógica muy particular que no especifica.

Si hablamos de intensiones es prudente hablar de Rudolf Carnap. Afirma Carnap, en la explicación que propone de los enunciados de creencia[3], que dos enunciados tienen la misma extensión si

---

[3] Cf. Carnap, R., *Meaning and Necessity*, Chicago, University, 1956 (1° Ed. 1947, sin apéndices). Apéndice C: "On Belief–Sentences" con réplicas de A.

ellos son equivalentes, esto es, si ambos son verdaderos o ambos son falsos. Por otro lado, dos enunciados tienen la misma intensión si ellos son lógicamente equivalentes, esto es, su equivalencia se debe a las reglas semánticas del lenguaje. Sea *A* un enunciado en que otro enunciado ocurre, digamos *p*. *A* se llama "extensional con respecto a *p*" si, y sólo si, la verdad de *A* no cambia si sustituimos el enunciado *p* con un enunciado equivalente *q*. *A* se llama "intensional con respecto a *p*" si, y sólo si:

1. *A* no es extensional con respecto a *p* y
2. La verdad de *A* no cambia si sustituimos el enunciado *p* con un enunciado lógicamente equivalente *q*.

Los siguientes ejemplos los debemos a Carnap[4]:

Primer ejemplo: El enunciado $(A \vee B)$ es extensional con respecto a *A* y con respecto a *B*; podemos sustituir *A* y *B* con enunciados equivalentes y el valor de verdad de $(A \vee B)$ no cambia.

Segundo ejemplo: Supongamos que *A* es **verdadero**, pero no **lógicamente verdadero**; por consiguiente los enunciados $(A \vee \sim A)$ y *A* son **equipolentes** (ambos son verdaderos) y, por supuesto, ellos no son **lógicamente equivalentes**. El enunciado $\Box(A \vee \sim A)$, donde $\Box$ es el operador de necesidad de la lógica modal, es verdadero y el enunciado $\Box A$ es falso; así $\Box A$ no es extensional con respecto a *A*. Al contrario, si *C* es un enunciado **lógicamente equivalente** a $(A \vee \sim A)$, entonces tanto $\Box(A \vee \sim A)$ como $\Box C$ son verdaderos: $\Box A$ es intensional con respecto a *A*.

Hay enunciados que no son ni extensionales, ni intencionales con respecto a otro; por ejemplo, los enunciados de creencia (como los suyos Prof. Bravo). El ejemplo de Carnap es "John cree que *D*". Supongamos que "John cree que *D*" es verdadero; sea *A* un enunciado equivalente a *D* y sea *B* un enunciado lógicamente equivalente a *D*. Es posible que el enunciado «"John cree que *A*" y "John cree

---

Church, p. 230-3. En el Apéndice D:. «Meaning and Synonymy in natural languages» Carnap defiende la "teoría de la intensión" ("teoría del significado", en términos de Quine) frente a las objeciones extensionalistas ("Teoría de la referencia") y presenta una concepción pragmática del significado lingüístico que, a su vez, intenta fundamentar empíricamente las nociones típicas de la teoría de la intensión, otorgándoles una explicación científicamente legítima.

4 *Ibid.*

que *B*'» sea falso. De hecho, John puede creer que un enunciado es verdadero, pero él puede creer que un enunciado lógicamente equivalente es falso. Para explicar los enunciados de creencia, Carnap define la noción de isomorfismo intensional. A grandes rasgos, dos enunciados son isomorfos intensionalmente si, y sólo si, sus elementos correspondientes son lógicamente equivalentes. En el enunciado de creencia "John cree que *D*" podemos sustituir *D* con un enunciado *C* intensionalmente isomorfo.

Si asumimos que *A* es el enunciado "Algunos perros son blancos" y *p* el enunciado "algunas cosas que son perros son blancas", como sugiere el Profesor Bravo, se tiene que mostrar, y ciertamente Bravo no lo hace:

a. Que *A* no es extensional con respecto a *p* y
b. Que la verdad de *A* no cambia si sustituimos el enunciado *p* con un enunciado lógicamente equivalente *q*.

Yo, honestamente, no veo cómo se pueda probar. A lo sumo parecen enunciados intensionalmente isomorfos. Valga la aclaratoria como un ejemplo de análisis de contenido intensional.

Bravo pretende que su análisis intensivo–descriptivo de los nombres permita un nuevo criterio ontológico para los lenguajes naturales. Para ello se basa en las opiniones de diversos lingüistas, incluyendo a Lyons. Pero Lyons no estaría de acuerdo con tal propuesta:

«…'Napoleón' se asocia arbitrariamente con muchas entidades (personas, animales, barcos, etc.), que no tienen en principio nada en común. […] una de estas entidades –o algún concepto, o comprensión, asociados con dicho nombre–, destaca en la cultura donde se usa normalmente el español por su importancia histórica […]. Esto significa que, a falta de información contextual específica en contra, para muchos hablantes de español, el nombre 'Napoleón' se considera normalmente como referido a la entidad culturalmente sobresaliente. También significa que habrá una gran cantidad de asociaciones compartidas y connotaciones agrupadas en torno al nombre 'Napoleón' que constituirán lo que muchos filósofos designan como la compresión, o concepto individual, "Napoleón". Sin embargo, *esto no significa que el nombre 'Napoleón' como tal tenga contenido o sentido descriptivo alguno*»[5].

En suma, se trata de la vieja tesis de Leibniz según la cual no hay términos del lenguaje que tengan núcleo semántico propio; esto es, que tengan un significado independiente de la teoría que los contenga.

---

5 Lyons, J.: *Semántica lingüística* [1995], Barcelona, Piados Ibérica, 1997, pp. 321-2.. (resaltado nuestro)

Bravo ha dejado palpable, gracias a lo que llamó "la disolución del sujeto óntico" (p.52), su carácter dogmático. Sí, dogmático. Ha tomado como incuestionable "la reducción radical" o "traducción radical" de enunciados a enunciados de experiencia inmediata, acaecimientos, datos sensibles, hechos atómicos, etc.; el llamado "segundo dogma del empirismo" y, con ello, su idéntico, el "dogma de la distinción analítico–sintético". No obstante, no dudo que encontrará también para los dogmas alguna "interpretación descriptiva".

Para Quine sólo las teorías sugieren lo que hay por medio del dominio de sus variables y no las "observaciones directas", ni lo "inmediatamente dado", ni "los hechos anteriores a toda interpretación teórica". Dice Quine:

«…insisto en que considero las variables y la cuantificación como evidencia de lo que una teoría dice que hay, no como evidencia acerca de lo que hay; pero este punto se pasa a veces por alto.[6]»… «Como empirista sigo concibiendo el esquema conceptual de la ciencia como un instrumento destinado en última instancia a predecir la experiencia futura a la luz de la experiencia pasada. *Introducimos con razón conceptualmente los objetos físicos en esta situación porque son intermediarios convenientes,* ***no por definición en términos de experiencia****, sino irreductiblemente puestos con un estatuto epistemológico comparable al de los dioses de Homero.* Yo por mi parte, como físico lego que soy, creo en los objetos físicos y no creo en los dioses de Homero, y considero un error científico orientar su creencia de otro modo. Pero en cuanto a fundamento epistemológico los objetos físicos y los dioses difieren sólo en grado, no en esencia. Ambas suertes de entidades integran nuestras concepciones sólo como elementos de cultura. El mito de los objetos físicos es epistemológicamente superior a muchos otros mitos porque ha probado ser más eficaz que ellos como procedimiento para elaborar una estructura más manejable en el flujo de la experiencia.»[7]

Desde el escorzo quineano las oraciones observacionales están representadas por los mismos asentimientos ante estímulos de la mayoría de los miembros de una comunidad de hablantes y esto es independiente de si el estímulo se reproduce sin la utilización de instrumentos de medición o con ellos. No hay preeminencia del aparato sensitivo humano como si fuera éste el único instrumento capaz de escrutar la referencia. Las observaciones que se producen con los distintos aparatos de medición, miden los "objetos", las "co-

---

6 Quine, W.V.O: *Palabra y objeto* [1960], Barcelona, Ed. Labor, 1968, p. 252 (nota al pie).

7 Quine: "Dos dogmas del empirismo" [1951] en *Desde un punto de vista lógico,* Barcelona, Ed. Orbis, 1984 (original inglés 1953), p. 79. (resaltado nuestro).

sas", que sugieren las teorías. Las teorías sugieren qué cosas hay. El criterio de Quine, el nivel de compromiso ontológico de las mismas nada tiene que ver con "su declarado nominalismo", como afirma Bravo en la página 44. ¿Dónde está el carácter normativo del criterio ontológico de Quine que Bravo clama? ¿Qué más descriptivo que la teoría de las descripciones definidas de Russell?

El lenguaje utilizado por Quine no es un lenguaje ni fenomenalista, ni fisicalista, ni reísta: si fuese un lenguaje fenomenalista o fisicalista no podría ejemplificar lo qué es una oración observacional hasta tanto no se reduzca una oración dada a su correspondiente oración equivalente dentro del lenguaje fenomenalista o fisicalista, pero tal reducción está excluida por la imposibilidad de la traducción radical; no es un lenguaje reísta porque no establece previamente qué tipos de cosas hay en el mundo, qué cosas corresponden a los términos del lenguaje o a las oraciones observacionales como denotata independientemente de la teoría que los contiene, pues tal posibilidad está excluida por la inescrutabilidad de la referencia intersubjetiva mediante significaciones estimulativas.

Quine plantea una alternativa ante lo que parecía una insuperable dicotomía entre lenguajes fenomenalistas o fisicalistas y reístas para el análisis epistemológico: la utilización del propio lenguaje de la ciencia con aquellas evidencias de lo que las teorías dicen que hay, todo ello conforme a su criterio de compromiso ontológico y a su planteamiento en "Dos dogmas del empirismo".

Instituto de Filosofía,
Universidad Central de Venezuela

EPISTEME NS, Vol. 20, Nº 1, 2000, pp. 135–140

JORGE NIKOLIĆ D.

# CREDO SEMÁNTICO DE UN INCONMENSURABILISTA COHERENTE

**Resumen:** Se muestra que el inconmensurabilismo no es una tesis histórica, sino una tesis semántica anclada en alguna teoría del significado (referencia–sentido) con consecuencias ontológicas, que se aplica a dos teorías científicas que intentan explicar la misma parcela del mundo, independientemente del momento histórico en que estas ocurran. Además, la tesis del inconmensurabilismo coherente con la afirmación que los objetos de nuestro estudio los *conocemos* o podemos llegar a *conocerlos*, es una incoherencia insalvable.

**Palabras Claves:** Inconmensurabilidad, referencia, sentido.

**Abstract:** In this article is explained that the incommensurabilism is not an historical thesis, but rather a semantic thesis anchored in some theories of the meaning (reference-sense) with ontological consequences. It applies as well to two scientific theories that try to explain the same piece of the world, whit no dependency on the historical moment that gives them place. In addition to this, the coherent incommensurabilism thesis, which asserts that *we do* know or that *we can get to* know the objects of our study, is an insurmountable incoherence.

**Keywords:** Incommensurabilism, reference, sense.

Desde la aparición del best–seller *La estructura de las revoluciones científicas –of course,* de Kuhn– ha corrido gran cantidad de tinta en las imprentas gracias al uso y abuso de dos términos: *paradigma* e *inconmensurabilidad.* El primero, *paradigma,* ya es de uso común, dependiendo su significado del usuario (ama de casa, político, boxeador, economista, etc...). El segundo, *inconmensurabilidad,* quizás debido a su fonética particular, no apta para disléxicos, esté mas confinado a los predios académicos.

Trivializada, la tesis de la *inconmensurabilidad*, permite a cualquier cagatintas académico proteger su coto de caza y enarbolar el *todo vale* sin límites; frente al cual, el inolvidable Fray Gerun-

dio, del Padre Isla, es un pobre aprendiz de brujo.

La susodicha tesis la barruntan más o menos así: las *teorías rivales* o *antagónicas,* las que están enfrentadas, no pueden ser reducibles entre sí a pesar de que intenten explicar lo mismo, es decir, no pueden compararse y, como también afirman que los significados de los términos cambian, entonces, cuando cambiamos de teoría decimos cosas diferentes. Luego, cada quién a fabricarse su propia torre sin posibilidad de comunicarse con el vecino, a menos que el vecino viva en la misma torre.

En general, la trivialización de la tesis, está bien rellena de psicología, sociología, religión, relaciones de poder, etc., con lo cual se concluye mediante el uso y abuso del *non–sequitur* otra perogrullada: que el comportamiento humano es humano. Nada nuevo bajo el sol.

Lo que se pretende en la líneas que siguen es precisar la tesis de la inconmensurabilidad mostrando la problemática filosófica asociada a ella, problemática que se retrotrae al viejo Aristóteles. Como el análisis que se va a realizar es semántico, aceptaré la siguiente premisa para todo el análisis posterior.

(F) El significado de un concepto viene determinado por el par: sentido, referencia, según Frege u otra versión mas sofisticada.

Entendiendo por referencia de un concepto lo que designa el concepto y por sentido al modo como el concepto designa su referencia. Si bien es cierto que no estamos aclarando mayormente lo que se entiende por referencia y sentido, también es cierto, que, nuestro propósito es mostrar la relación de dependencia de las tesis de inconmensurabilidad con algún aparato semántico y alguna ontología, por lo tanto una discusión exhaustiva acerca del sentido y la referencia es por los momentos irrelevante.

*1.- La tesis de la inconmensurabilidad trivial.*

Dicha tesis afirma:

- la.- La premisa semántica (F).
- lb.- Existe una relación metateórica denominada inconmensurabilidad trivial que se da entre dos teorías científicas cualesquiera, que no explican la misma parcela del mundo.
- lc.- Los términos de una teoría no pueden reconstruirse semánticamente unos a partir de los otros.
- ld.- Es imposible la traducción de enunciados entre las dos teo-

rías (consecuencia de lc).

le.- No existe una teoría neutral respecto a la cual las dos teorías serían traducibles.

lf.- Los sentidos y las referencias de los términos de ambas teorías son diferentes.

Es trivial que teorías con parcelas distintas del mundo tengan términos distintos con semánticas distintas y no sean traducibles; por ejemplo: la teoría del electrón de Dirac y la teoría de la conducta sexual de los pingüinos. Este caso de inconmensurabilidad no es de mayor interés, de allí la etiqueta de trivial.

*2.- La tesis de la inconmensurabilidad no–trivial (o inconmensurabilidad comparable) relativa al sentido.*

La tesis afirma :

2a.- La premisa semántica ( F )

2b.- Existe una relación metateórica denominada inconmensurabilidad comparable, o no–trivial relativa al sentido, que se da entre dos teorías científicas que pretenden explicar la misma parcela del mundo, –teorías conocidas como *rivales* en el argot científico–filosófico.

2c.- Los términos primitivos de dos teorías rivales -a pesar de que se escriban y pronuncien igual y tengan algunos de ellos sintaxis muy parecidas o iguales- no pueden reconstruirse semánticamente unos a partir de los otros.

2d.- Es imposible la traducción de enunciados entre las dos teorías rivales ( consecuencia de 2c ).

2e.- No existe una teoría neutral respecto a la cual las dos teorías serían traducibles.

2f.- Los sentidos de los términos de uso común en ambas teorías son diferentes pero sus referencias son iguales, ya que existen procedimientos para determinarlas, que no varían al cambiar las teorías

El antecedente histórico de 2e. se remonta a Locke y su afirmación acerca de la carencia de un núcleo semántico fijo por parte de los términos, es decir, su significado es contextual.

Manteniéndose dicha postura en la concepción heredada o carnapiana. Es decir, si añadimos, quitamos o modificamos al menos un axioma a la teoría, automáticamente se da un cambio en el significado de los términos, dándose la divergencia radical de signi-

ficado. Y como la traducción de enunciados requiere de la permanencia del significado, la cual al no darse, hace imposible la traducción.

Si existiera una teoría neutral respecto a la cual las dos teorías serían traducibles, entonces, debido a la simetría y transitividad de la traducción, las dos teorías rivales serían traducibles violando 2d.

Si nuestro inconmensurabilista decidiese identificar la referencia de los términos científicos con los estímulos sensoriales y su aparato lógico–matemático, pasaría a ser un fenomenalista tipo *Aufbau* o un empirista lógico siempre y cuando admitiese a los términos teóricos. Y si tomase postura por un realismo podría perfectamente adoptar la tesis de un realismo referencial 2f. ontológico y epistemológico : hay cosas independientes de mí en el mundo, a los que se refieren los conceptos de las teorías y que además en algunos casos, puedo saber cuales son esas cosas.

*3.- La tesis de la inconmensurabilidad no–trivial (ó inconmensurabilidad comparable) relativa al sentido ya la referencia, o inconmensurabilidad referencial.*

La tesis afirma: (3a), (3b), (3c), (3d), (3e) lo mismo que en (2a), (2b), (2c), (2d), (2e), además :

*3f.- Los sentidos y las referencias de los términos de ambas teorías son diferentes.*

La justificación de 3f. en especial lo que concierne a la referencia lo haremos en el punto siguiente, el 4. Existen justificaciones de 3f basadas en interpretaciones históricas, todas ellas harto problemáticas y que el lector conseguirá fácilmente. La discusión de 3 la remitimos a 4.

*4.- La tesis de la inconmensurabilidad quineana.*

La tesis afirma: 4a, 4b, 4c, 4d, 4e, 4f lo mismo que en 3a, 3b, 3c, 3d, 3e, 3f, además :

4g.- El principio de la inescrutabilidad de la referencia de Quine. La referencia de un término es relativa a su relación con otros términos dentro de cierto marco lingüístico, por lo tanto, no tiene sentido tomar la referencia como un absoluto.

4h.- No existe una teoría científica (una superteoría referencial), que sea un marco referencial universalmente admitido para todas las teorías existentes.

La argumentación es idéntica a la que aparece después de 2f. con excepción de lo que se refiere al realismo. Aclaremos 4g. mediante un ejemplo. La referencia de *conejo,* solo puede conocerse si conocemos las relaciones entre los referentes de los términos relacionados con *conejo,* como *zanahoria, orejas largas, saltar,* etc. Además, la identidad de referencia del término *conejo* al ser usado por dos individuos que usen el mismo lenguaje, es una hipótesis empírica que nunca podrá ser comprobada. Tampoco podemos afirmar dicha identidad al ser usado el término por el mismo individuo en dos instantes diferentes.

Si existe la teoría postulada en 4h manteniéndose 4g, se refutaría la tesis de la inconmensurabilidad, ya que la referencia de los términos de todas la teorías estaría determinada por la teoría postulada en 4h.

La diferencia ontológica de esta tesis con la planteada en 2 estriba en la postura realista. Como la tesis 4 nos afirma que no podemos conocer la referencia de los términos de una teoría dada cualquiera, entonces, si afirmamos ser realistas, no podemos serlo epistemológicamente sino sólo ontológicamente, es decir, tenemos que renunciar al conocimiento de los objetos que designan las referencias de dichos objetos. ¿Entonces, qué le queda por hacer al realista? Encontrar las condiciones que han de darse para la *existencia* de dichos objetos, y que podremos *conocerlos* valga la metáfora, sólo de modo *aproximado*.

Pero intentar validar la *existencia* de dichos objetos con independencia de individuos y teorías sería entrar en la problemática argumentación de la teoría causal de la referencia de Putnam y Kripke. La otra postura realista sería la del realismo alético a la Quine con el inconveniente de ser un realismo *poco realista* ya que la noción de verdad absoluta no tiene relevancia epistemológica.

En resumen, si nuestro inconmensurabilista es coherente :

a.- Tiene su tesis anclada en una teoría del significado, no hacerlo o negarlo, es incurrir en dislates.

b.- La ontología que se deriva depende de las restricciones semánticas a elegir.

c.- Los realismos no quedan bien parados al aceptar la tesis de la inconmensurabilidad quineana.

Ser inconmensurabilista coherente y andar pregonando que *conocemos* o podemos llegar a *conocer* a los objetos de nuestro es-

tudio es una incoherencia insalvable

d.- El inconmensurabilismo no es una tesis histórica, es una tesis semántica, con consecuencias ontológicas, que se aplica a dos teorías científicas que intentan explicar la misma parcela del mundo, independientemente del momento histórico en que estas ocurran.

Instituto de Filosofía,
Universidad Central de Venezuela

## BIBLIOGRAFÍA

Moulines, U., *Pluralidad y recursión,* Madrid, Ed. Alianza, 1992.

Quine, W., . *La relatividad ontológica y otros ensayos*, Madrid, 1974.

JORGE NIKOLIĆ D.

# LA IMPOSIBILIDAD DE EVITAR A LA FILOSOFÍA

**Resumen:** A partir de la reflexividad recursiva de cualquier discurso y de la demarcación del trabajo filosófico como reflexión en segunda o mayor potencia se motiva la premisa: no existe ítem de información ni discurso que no presuponga al menos a alguna premisa filosófica.

**Palabras Claves:** Reflexividad recursiva, metalenguaje.

**Abstract:** The premise: any item of information or of speech presupposes, at least, one philosophical assumption, is motivated by the recurrent reflexivity of any speech and by the philosophical work which is demarcated as a reflection of a second or of a greater power.

**Keywords:** Recurrent reflexivility, metalanguage.

Es harto conocido que el quehacer filosófico desde una óptica analítica consiste de reflexiones en segunda o mayor potencia, es decir, si se asume la premisa que afirma*: podemos pensar acerca de lo que pensamos*, concluimos que todo discurso es reflexivamente recursivo. Lo anterior apunta a la demarcación del trabajo filosófico, labor que no tiene porque interesar necesariamente a los especialistas dedicados a dar cuenta de una determinada parcela del mundo mediante –al menos– un lenguaje.

Luego, un especialista –no filósofo–, puede argumentar que mientras su interés no radique en reflexionar acerca de su lenguaje, la filosofía le es irrelevante e innecesaria, incluso en las acciones vitales de su vida cotidiana cuando se comunica usando el lenguaje natural. Muchos predican de la artificiosa pedantería del quehacer filosófico, afirmando que dicha labor está divorciada del trabajo científico en lo que se refiere a la búsqueda de soluciones a determinados problemas, a la redacción de informes y textos, además de su inutilidad en el contexto de descubrimiento.

A continuación vamos a motivar la siguiente hipótesis: *no existe ítem de información ni discurso –por simple que sea– que no presuponga al menos a una metapremisa filosófica.* Supondremos que dicho discurso tiene pretensiones de coherencia, lo que equivale a afirmar que no podemos decir *lo que se nos venga en gana,* con lo cual intentamos excluir a las contradicciones.

Imaginemos primero a una madre regañando a su hijo (lo cual no es difícil) que tiene un dolor de estómago.

-¡Tienes dolor de estómago porque comiste alguna porquería en la calle!– dice la madre, arremetiendo a seguidas con: –eso te pasa por no hacerle caso a tu madre, que es la única que te canta las verdades, merecido tienes tu dolor!..., etc. En primer lugar, la madre apela al principio de causa–efecto; el dolor de estómago es un efecto, luego ha de haber una causa: comer porquerías. Ella afirma que existe la *verdad,* ella la posee, y todo lo que ella diga, lleva ese sello, y si su hijo sigue la guía de sus consejos, que al parecer son todos verdaderos, no le pasará nada malo y tendrá éxito.

Imaginemos ahora a un físico explicando la caída de un pequeño objeto pesado (lo cual tampoco es difícil de imaginar). –La piedra que tenemos en la mano cae al soltarla, debido a que sobre ella actúa la fuerza de la gravedad–, afirma nuestro físico, para luego agregar, –lo cual es un enunciado comprobado en todos los casos, no hay dudas acerca de ello, es una *verdad.* Nuestro docto físico apela –lo mismo que la madre– al principio de causa–efecto, la piedra cae –efecto–, debido a que cesó la fuerza que la mantenía en alto cuando la soltamos, y a la fuerza de la gravedad, –causa. Pero nuestro físico no es como la madre de nuestro ejemplo, ya que no se atreve a afirmar que él posee la verdad o todo lo que él diga es verdadero; sino que apela a todo el colectivo de físicos, los especialistas.

El físico dice que sus colegas han comprobado la existencia de la gravitación y sus regularidades, luego la ley de la gravitación es *verdadera.* Además, nos dice: para explicar correctamente cualquier caída de un cuerpo pesado, y poder hacer predicciones, tendremos que apelar a la ley y seguir los preceptos de la mecánica.
Tanto a la madre como a nuestro docto físico, los pondríamos en calzas prietas si les preguntamos:

i.- ¿Qué es la relación causa–efecto?
ii.- ¿Qué es la *verdad* y como se predica acerca de ella?
iii.-¿Por qué en ambos casos los enunciados que hemos de aceptar

para tomar decisiones supuestamente racionales a nivel de actividades vitales de los humanos, tienen que ser *verdaderos*?

Ni el lenguaje de la madre ni las teorías físicas del físico contienen las respuestas a las tres últimas preguntas, pero las presuponen en su discurso, peor aún, no sabrían transmitir la información que dieron si les prohibieran el uso de dichas *metapremisas.* Cuando emitimos alguna información suponemos que tiene algún *significado.* Cuando explicamos asumimos que existen las *explicaciones*. Afirmamos, negamos, normamos, describimos e inferimos, suponiendo que es posible *afirmar*, *negar*, *normar, describir* e *inferir.* Pero a pesar de que nuestro discurso no trate acerca del significado, la explicación, la verdad o la consecuencia, no puede prescindir de ellas. Es difícil imaginar algún discurso en el que no aparezcan. No hemos demostrado nada, pero hemos sugerido que no existen lenguajes que no contengan premisas filosóficas –seguramente, muchas. Lo anterior motiva la siguiente *metapremisa* filosófica:

*Todo lenguaje presupone alguna metapremisa filosófica o de modo mas preciso, toda teoría contiene al menos una metapremisa filosófica.*

Cada vez que decimos algo deseando comunicarnos suponemos la *metapremisa* de la *significatividad* de nuestro discurso o que hay una relación entre lo que decimos y algunas entidades translingüisticas.

Cada vez que suponemos la existencia de objetos reales independientes del sujeto cognoscente dando cuenta de ellos con algún lenguaje tenemos varias *metapremisas* filosóficas. Es decir, si apunto con mi dedo a un lápiz y digo: "Esto es un lápiz", estaría consciente o inconscientemente suponiendo alguna noción de significado y en consecuencia, algún tipo de realismo.

Cuando argumentamos inductiva o deductivamente aceptando y rechazando ítems de información, cuando hablamos de conceptos y de leyes, cuando explicamos o hacemos prognosis, cuando normamos o describimos, o simplemente cuando usamos sin cuestionar alguna metodología sugerente para resolver algún problema, estamos suponiendo premisas filosóficas. Más aún, exigir de un discurso que sea objetivo es aceptar una premisa filosófica: la objetividad.

Cuando en un contexto de descubrimiento normamos las propiedades que ha de tener una teoría estamos hablando de *metapropiedades,* es decir, propiedades de propiedades; dichas exigencias

son premisas filosóficas. Basta revisar desde la noción de cambio en los presocráticos pasando por el criterio de objetividad de Galileo, hasta llegar a las modernas teorías de invariantes que se usan en las teorías físicas actuales.

Mientras más primitivo es el estado del desarrollo de una teoría se hace más evidente la apelación a premisas filosóficas. En las teorías sofisticadas y exitosas el enramado de *metapremisas* es más denso, gracias a ello el especialista no se da por enterado, quizás en parte debido al modo acrítico de usar su metalenguaje, siendo esta, quizás su característica principal desde una óptica filosófica, –*of course*. Además, en caso de aceptar a alguna metapremisa la transforma en un principio de fe que pertenece a su disciplina.

En resumen, cada vez que suponemos algo que pertenece al sentido común o que no cuestionamos debido a considerarlo parte de nuestros *chips* genéticos, de seguro estamos en presencia de al menos una premisa filosófica.

Las más simples y evidentes son: las que aceptamos cuando usamos nuestro lenguaje al intentar comunicarnos. Explicamos, significamos, justificamos, aseveramos y hasta nos molestamos cuando alguien nos miente y, suponemos que *significar, explicar, justificar, normar, describir,* o el manejo del término *verdad* es algo dado, que se adquiere por el uso, pero que no se cuestiona.

La conclusión es evidente, si no podemos evitar las metapremisas de cualquier discurso o ítem de información a pesar de que no estemos conscientes de ello, no podemos evitar a la filosofía.

La moraleja es aún más evidente: si como aquel señor *Jourdain* quien aprendió a hablar en prosa, nos hemos dado cuenta que cada vez que hablamos no podemos evitar a la Filosofía, podríamos ir más lejos que *Monsieur Jourdain.* Como las *metapremisas* son las que definen las fronteras del lenguaje, aprender su uso equivale a llegar a conocer los alcances y los límites de nuestros lenguajes.

Instituto de Filosofía,
Universidad Central de Venezuela

CHOMSKY, N.: *Language and Thought,* Rhode Island & London, Moyer Bell Ed., 1993, pp. 96.

El texto recopila una conferencia dictada por Noam Chomsky, y la subsecuente discusión, en ocasión de la tercera Lectura Transdisciplinaria en Arte, Ciencia y Filosofía de la Cultura dirigida por Ruth Nanda–Anshen de la Real Sociedad de Artes de Londres.

Chomsky presenta un claro resumen de su concepción de la relación de la lingüística con el estudio de la mente, de la relación entre lenguaje y pensamiento. A partir de la consideración de lo que llama la representación de Frege de la relación entre lenguaje y pensamiento, rechaza varias preguntas típicas de los filósofos y los científicos cognitivos, como la pregunta ¿pueden pensar las máquinas?, por estar mal planteadas y, por consiguiente, no las considera dignas de un análisis particular hasta una subsecuente aclaración. Cuestiona, igualmente, la prueba del cuarto Chino propuesta por Searle. También encuentra el llamado 'naturalismo metafísico', en contraste con el 'naturalismo metodológico' que él profesa, como una postura filosófica dogmática y, en algún sentido, incoherente. Después de todas estas críticas, ofrece algunas sugerencias positivas sobre la relación entre lenguaje y pensamiento. Y, en medio de toda esta compleja y detallada exposición, proporciona un diagnóstico histórico bastante interesante de lo que considera correcto e incorrecto en la filosofía cartesiana, y censura la moda de este siglo que trata a Descartes como un leproso filosófico. Desde su punto de vista, Descartes planteó la primera revolución cognitiva con su teoría mente–cuerpo.

Para rechazar como una pregunta mal planteada ¿Pueden pensar las máquinas? parte de la pregunta ¿confirmamos o refutamos el Test de Turing considerando la posibilidad de una máquina que reproduce nuestra conducta finita? Una respuesta la esboza considerando un problema análogo. Considérese un sistema típico de entradas y salidas, la respiración. Toscamente hablando, lo que ocurre es que el aire entra en la nariz y expulsamos dióxido de carbono. Supóngase que se puede conseguir una máquina que duplique completamente tal sistema por medio de algunos mecanismos. ¿Estaría respirando la máquina? La máquina no estaría respirando por razones triviales. Respirar es una cosa que hacen los humanos, por consiguiente, la máquina no está respirando. ¿Es un buen modelo de humanos? Para dar una respuesta, Chomsky indica que observaría si tal máquina nos enseña algo sobre los humanos. Si lo hace, es un buen modelo de humanos. Si no enseña algo sobre los humanos, Chomsky la enviaría a las llamas de Hume.

A Chomsky le parece que tal planteamiento se aplica exactamente a la relación entre pensamiento y lenguaje. Alega que alguien podría venir con un programa de ajedrez que hace exactamente los mismos movimientos que cada vez haría Kasparov. Se pregunta ¿La máquina estaría jugando ajedrez? No; tal como en el caso de la respiración. Jugar ajedrez es algo que

hacen las personas. Kasparov tiene un cerebro, pero su cerebro no juega ajedrez. Si se pregunta, "¿El cerebro de Kasparov juega ajedrez?", la respuesta es no; lo contrario sería como afirmar que sus piernas toman un paseo. Según Chomsky, es un punto trivial; no es un punto interesante para discutir. Chomsky indica que nuestras piernas no toman un paseo, nuestro cerebro no juega ajedrez o entiende inglés. Simplemente por la misma razón que un submarino no nada. Nadar es algo que hacen los peces. Si se quiere extender la metáfora a los submarinos, podríamos decir que ellos lo hacen. Con respecto al inglés escoge una metáfora diferente, pero éstas preguntas no son substantivas. Una máquina que reproduce el intercambio de aire a dióxido de carbono no estaría respirando por razones triviales, así como un robot que clava un cuchillo a alguien en el corazón, no estaría asesinando. Los robots no pueden asesinar. Eso es algo que hacen los humanos. Por estas razones, tales preguntas no significan nada.

En suma, Chomsky puede no tener la razón, pero el texto es un agradable llamado a la cordura con respecto al tratamiento del problema lenguaje–pensamiento.

JESÚS F. BACETA V.
Universidad Central de Venezuela
Facultad de Humanidades y Educación
Instituto de Filosofía

---

Ilham Dilman, *Free will: a historical and philosophical introduction*, New York, Routledge, 1999, p. 266.

El problema del libre albedrío es un problema con un largo historial en la historia de las ideas de Occidente. Ya desde la misma Grecia era un asunto que ocupaba a Filósofos y Literatos, y hasta nuestros días sigue siendo un problema que es motivo de incesante discusión en círculos académicos y jurídicos. No es de extrañarse que el tema tenga tal repercusión, no sólo desde el punto de vista académico, sino también desde el punto de vista de las interacciones sociales en la vida cotidiana. Las interacciones de los seres humanos y sus instituciones estarán inexorablemente condicionadas por la opinión que se tenga sobre el libre albedrío. A partir de esta idea se atribuye responsabilidad a una persona por cierta acción, se exonera de culpa, se evalúan los hechos pasados y la posibilidad de lo que se pueda hacer en el futuro. En fin, se pone de manifiesto todo aquello por lo cual el ser humano es, o pretende ser, un ser racional. De ahí que el problema sea de no poca relevancia.

En su texto, *Free will: a historical and philosophical introduction,* Ilham Dilman presenta una introducción a la problemática y ciertos autores

fundamentales que sirven para entender el desarrollo del tema en la historia de Occidente. El problema es sumamente complejo y ha tenido múltiples facetas, de manera tal, que hablar del tema en nuestros días dista mucho de la naturaleza del problema en la antigua Grecia o en la Edad Media. De hecho, al examinar la naturaleza del problema en diferentes momentos históricos nos daremos cuenta que realmente se trata, no sólo de diferentes facetas del problema, sino más bien de diferentes problemas propiamente. Sin embargo, todos ellos tienen algo en común: el tema de la libertad humana. En este sentido el texto de Dilman presenta una muy buena delineación para entender la naturaleza del problema, o mejor dicho de los problemas, y su continuidad histórica. A grandes rasgos el tema tiene tres momentos históricos y tres dimensiones diferentes respectivamente. El problema de un destino determinado en la antigua Grecia, el problema de la determinación en la Salvación desde el punto de vista de la teología Cristiana, y el problema del determinismo natural en el mundo Científico y Moderno. Hay una sección dedicada a cada uno de estos temas, de forma tal, que los capítulos del libro están presentados de manera cronológica y están agrupados de acuerdo a las divisiones de estos tres momentos.

La primera sección del texto tiene como título: "Early Greek thinkers, moral determinism and individual responsibility". En ella se analizan *La Iliada* de Homero, el *Edipo Rey* de Sófocles, varios textos de Platón, especialmente el *Gorgias*, y finalmente los libros III y VI de la *Ética a Nicómano* de Aristóteles. El problema en este grupo de autores, de acuerdo con el análisis de Dilman es el siguiente: "As far as the problem of human freedom goes their man concern in the way human beings become the plaything of certain common human propensities" (pl.).

En una segunda parte del libro Dilman se encarga del problema específicamente teológico. En esta segunda parte, la cual lleva por título "The coming of age of Christianity: morality, theology and free will", se analizan textos de San Agustín y de Santo Tomás de Aquino. El texto relevante, a partir del cual Dilman discute el problema del libre albedrío en San Agustín es *De libero arbitrio*, y en el caso de Santo Tomás el texto relevante es *De veritate*. En esta sección se tratan temas que tienen que ver con la omnisciencia de un Ser todopoderoso, la predestinación, y la manera en que para los autores medievales estos problemas fueron una constante fuente de reflexión. El punto central de toda la reflexión viene a lo siguiente: conciliar el carácter necesario de un acontecimiento, puesto que de alguna manera la omnisciencia del Todopoderoso lo determina, y el libre albedrío de los mortales. El argumento es a grandes rasgos el siguiente: si Dios conoce todo lo que acaecerá entonces todo lo que ocurre debe ocurrir *necesariamente*, si todo lo que acaece, ocurre de manera necesaria, entonces no hay espacio para la contingencia y mucho menos para el libre albedrío de los seres humanos. Dicho de otra manera, las cosas no pueden o pudieron haber sido diferentes a como efectivamente fueron. La gravedad del problema consiste en que si los eventos ocurren de manera necesaria entonces

los seres humanos no pueden ser moralmente responsables de sus vidas y esto convierte a Dios, o bien en un juez sumamente injusto, o en el autor y sustentor del Mal. Este fue un tema muy prolijo en el Medioevo que también vino a tener mucha importancia durante la reforma protestante.

Otra faceta del problema es la relación que existe entre las leyes causales del universo y la posible contingencia de los acontecimientos. Una vez secularizado el problema de la predestinación, la naturaleza de la posible determinación de los acontecimientos también muda de atuendos. El problema ya no tiene que ver con un Ser transcendental que posiblemente determine los acontecimientos debido a su omnisciencia, sino que el contrapunto es ahora realizado con la idea de Naturaleza. El ser humano es entendido como miembro de la Naturaleza y por lo tanto propenso a sus mismas características, entre las cuales se encuentran la regularidad, la nomología y la causalidad. El problema pasa ahora a tener la siguiente forma: si a todo acontecimiento le antecede una causa de acuerdo a cierta regularidad en el universo, entonces todo acontecimiento es causalmente determinado, y de ser así, cómo puede existir el libre albedrío. Lo que se pone en cuestión, en todo caso, es la agencia de los seres humanos y su carácter moral.

De esta faceta del problema se ocupa la tercera parte del libro de Dilman, en una sección que lleva por título, "The rise of science: universal causation an human agency". En esta sección se discuten las *Meditaciones* de Descartes, la *Ética* de Spinoza y diversos textos de Hume y Kant, entre los cuales se encuentran, *An enquiry concerning the principles of morals*, y la *Fundamentación de la metafísica de las costumbres*. De particular importancia para estos autores es el problema de los seres humanos enmarcados en un universo natural que les determina causalmente y en el cual debe encontrarse espacio para el libre albedrío. De no ser así, no puede existir tal cosa como la expectativa y retribución moral en los seres humanos.

La cuarta y última sección del texto es realmente una extensión de la tercera. La única diferencia es que el problema se vuelve mucho más sofisticado debido a que, por un lado se discute y quiere justificar, o poner en cuestión, el carácter esencialmente libre de los seres humanos, y por otro lado, se trata de conciliar la idea del libre albedrío con la naturaleza mecanicista del universo, tal y como lo proponen las ciencias físicas. Las conclusiones de los autores son sumamente variadas y cabe decir que no siempre optimistas. Sencillamente, puede llegar a asumirse una fatalidad existencial en la cual la misma libertad se convierte en una forma de determinismo, Entre los autores para quienes el punto de la libertad y la posible determinación no sólo de la causalidad natural sino también de los deseos e irracionalidad humana, es sumamente importante, encontramos a Schopenhauer, Freud Sartre y Simone Weil. Por otro lado, entre los autores que dialogan expresamente con el paradigma cientificista, encontramos a Moore y Wittgenstein. Los textos de cada autor discutido en esta sección

son los siguientes: Schopenhauer, *On the freedom of de will*[1]; Sartre, *El ser y la nada*; Freud, *Más allá del principio del placer*; Simone Wein, *La pensateur et la grâce*; G.E, Moore, *Ethics*; Wittgenstein, *Lecture on the freedom of the will*. Esta última sección del libro de Dilman lleva por título: *The age of psychology: reason and feeling, causality and free will.*

La mayor virtud del texto en cuestión es que provee un amplio arqueo histórico del problema del libre albedrío. Así, se hace mucho más claro el entender y precisar cual es la naturaleza e importancia del problema. Dilman no se limita a exponer los autores en cada sección del libro, sino que también dialoga y hasta muestra abiertamente sus preferencias en el asunto. Sus dos autores predilectos, y de alguna manera desde los cuales lee y juzga a los demás autores, son Weil y sobre todo Wittgenstein. El texto de Dilman es entonces una introducción al problema del libre albedrío desde el punto de vista primordialmente histórico y tangencialmente filosófico. El propósito y el espíritu del texto está bien expresado en su último parágrafo: "The book thus considers the contribution of a number of thinkers in Western thought from the time of Ancient Greece to the present (…). It brings out the richness of what is in question and itself contributes to the discussion of the questions raised by these thinkers under its different aspects."(p.266)

JOSÉ E. IDLER
Estudiante de la Maestría en Filosofía y Ciencias Humanas
Universidad Central de Venezuela

1 He listado este texto en Inglés, puesto que Dilman lo trabaja en ese idioma. Ignoro el original en alemán o el nombre del texto en su traducción al español.

μ

Los profesores Omar Astorga, Luz Marina Barreto, Daniel Hernández, Carlos Kohn y Nancy Núñez participarán en las *III Jornadas de Análisis del Discurso Político,* dentro del marco del *III Coloquio Nacional de Análisis del Discurso*, el cual está siendo organizado por la Asociación Latinoamericana de Estudios del Discurso, Delegación Regional (ALED), la Universidad Nacional Experimental "Francisco de Miranda" (UNEFM), el Instituto Universitario de Tecnología "Alonso Gamero" (IUTAG), Fundacit-Falcón y la Comisión Regional para el Mejoramiento de la Enseñanza de la Lengua Escrita (CORMELE). Dichos eventos se celebrarán en la ciudad de Santa Ana de Coro, del 28 al 30 de septiembre de 2000, en las instalaciones de la Universidad "Francisco de Miranda".

μ

La Asociación Venezolana de Filosofía Política, con el coauspicio del Consejo de Desarrollo Científico y Humanístico de la U.C.V, Postgrado de la Facultad de Humanidades y Educación de la U.C.V, Postgrado en Filosofía de la Universidad de los Andes, Instituto de Filosofía del Derecho de la Universidad del Zulia y del Postgrado de la Facultad de Ciencias Jurídicas Políticas de la U.C.V, celebraron el *Primer Coloquio Internacional de Filosofía Política* sobre Crisis, Gobernabilidad y Proyecto Político Venezolano, el cual se realizó del 5 al 7 de julio del 2000 en la Sala de Usos Múltiples del Postgrado de Humanidades y Educación de la U.C.V.

μ

Durante el mes de junio del presente año se celebrará en la ciudad de Salamanca, España, el *First International Congress for Teaching Logic*, organizado por la Universidad de Salamanca. Parte de la temática del Congreso se centrará en la presentación de software educativo especialmente orientado a la enseñanza de la lógica. Entre los miembros del Comité Organizador tenemos a la Prof. María Manzano de la Universidad de Salamanca, el Prof. Dick de Jongh de la Universidad de Amsterdam, así como a distinguidos investigadores del grupo ARACNE. Por Venezuela se espera la presencia la presencia de los profesores Ezra Heymann, Tulio Olmos y Levis Zerpa de la Universidad Central de Venezuela y de la Prof. Corina Yoris de la Universidad Católica "Andrés Bello". Para mayor información al respecto consultar la página Web: http://aracne.usal..es/congreso/congress.html.

μ

El Prof. Carlos Kohn, miembro del Instituto de Filosofía ha sido nombrado representante de la Facultad de Humanidades y Educación en el Comité Editorial de la Revista *Tharsis,* nuevo vocero académico de las Facultades de Humanidades, FACES y Ciencias Jurídicas y Políticas de la UCV.

μ

La Revista Internacional de Filosofía Política, la Universidad de Antioquía y la Universidad de Cartagena de Indias están organizando el VII Simposio de la Revista Internacional de Filosofía Política, el cual se celebrará en la ciudad de Cartagena, en los días de noviembre del 2000.

μ

El Decano de la Facultad de Humanidades y Educación de la UCV está invitando a la Exposición de las ediciones de la Facultad, que se realizará en los espacios de la librería Monte Ávila, en el Complejo Cultural "Teresa Carreño, del 27 de junio al 9 de julio de 2000.

# LIBROS RECIBIDOS

Alchourrón, Carlos (ed.). *Lógica.*
Enciclopedia Iberoamericana de filosofía. Editorial Trotta. Madrid, 1995.
366 pp.

Contenido. 1. Introducción: Concepciones de la lógica. 2. Lógica clásica de primer orden. 3. Lógica de orden superior. 4. Lógica deóntica. 5. Lógica e inteligencia artificial. 6. Lógica para consistente. 7. Lógica epistémica. 8. Lógica temporal. 9. Lógica cuántica. 10. Lógica de la relevancia. 11. Computabilidad. 12. Lógica modal. 13. Lógicas multivalentes.

Barwise, Jon; Etchemendy, John. Language proof and logic.
CSLI Publications, Stanford, 1999.587 pp.

*Contenido*. 1. Introduction. 2. Propositional logic: Atomic sentences; The logic of atomic sentences; The boolean connectives; The logic of boolean connectives; Methods of proof for boolea connectives; Formal proofs and boolean logic; Conditionals. 3. Quantifiers: Introduction to quantification; The logic of quantifiers; Multiple quantifiers; Methods of proof for quantifiers; Formal proofs and quantifiers; Moere about quantification. 4. Applications and metatheory: Frst-order set theory; Mathematical induction; Advanced topics in propositional logic; Advanced topics in FOL; Completeness and incompleteness; Summary of formal proof rules.

Camps, Victoria: *Qué hay que enseñarle a los hijos.*
Ed. Plaza Janés, Barcelona, 2000.
181 pp

*Contenido:* 1. Todos somos hijos, por Margarita Rivière. 2. Prólogo. 3. Felicidad. 4. Buen Humor. 5. Carácter. 6. Responsabilidad. 7. Dolor. 8. Autoestima. 9. Buenos sentimientos. 10. Buen gusto. 11.Valentía. 12. Generosidad. 13.Amabilidad. 14.Respeto. 15. Gratitud. 16. Hijos/Hijas. 17. Trabajo. 18. Televisión. 19. Libertad. 20. Obediencia. 21. Ejemplo y tiempo. 22. Saber más: Antología y lecturas.

Casalta, Henry: *Reflexiones sobre temas y conceptos del análisis conductual.*
Ed. Comisión de Estudios de Postgrado, Facultad de Humanidades y Educación, UCV, Caracas, 1999.
94 pp.

*Contenido:* Presentación. 1. Evaluación actual del enfoque de B.F Skinner: alcance y limitaciones. 2.Claves para la comprensión de la Conducta verbal. 3. El concepto de repertorio en el análisis de la Conducta Verbal.3. El tiempo como "precepto" y como concepto: Sugerencias para su investigación empírica.

Casalta, Henry y Lacasella, Rosa: *Compendio de la conducta verbal de B.F. Skinner.*
Ed.Comisión de Estudios de Postgrado, Facultad de Humanidades y Educación UCV, Caracas, 1999.
84 pp.

*Contenido:* Presentación. 1. Un análisis funcional de la conducta verbal. 2. Problemas generales. 3. Variables controladoras. 4. La conducta verbal bajo el control de los estímulos verbales. 5. El tacto. 6.Condiciones especiales que afectan al control por los estímulos. 7. La audiencia. 8. La operante verbal como unidad de análisis. 9. Causación múltiple (múltiples variables). 10. Estimulación suplementaria. 11. Nueva combinación de respuestas fragmentarias. 12. La mnipulación de la conducta verbal. 13. Gramática y sintáxis como procesos autoclíticos. 14. La composición y sus efectos. 15.Autocorreción. 16. Condiciones especiales de autocorrección. 17. El autorreforzamiento de la conducta verbal. 18. Conducta verbal lógica y científica. 19. Pensamiento. Glosario de algunos términos técnicos de la gramática y la retórica. Temas y sugerencias para ser investigados.

Davidson, Donald: *Ensayos sobre acciones y sucesos.*
Ed. Crítica, Barcelona, 1995.
382 pp.

*Contenido:* Introducción. Intención y acción. 1. Acciones, razones y causas (1963). 2. ¿Cómo es posible la debilidad de la voluntad? (1970) 3. De la acción (1971). 4. Libertad para actuar (1973). 5. Tener la intención (1978). Suceso y causa 6. La forma lógica de las oraciones de acción (1967) Críticas comentarios y defensa. 7.Relaciones causales (1967). 8. La individuación de los sucesos (1969). 9. Los sucesos como particulares (1970). 10. Sucesos eternos vs sucesos efímeros (1971). Filosofía de la psicología. 11. Sucesos mentales (1970) Apéndice: esmerrosas con otros nombres (1966). 12. La psicología como

filosofía (1974) Comentarios y respuestas. 13. La mente material (1973). 14. Hempel y la explicación de la acción (1976). 15. La teoría cognoscitiva del orgullo de Hume (1976). Bibliografía. Índice analítico.

Lo Monaco, V.P., *La nueva metafísica de la lógica modal,*
Editorial de la Comisión de Estudios de Postgrado. Caracas.1999.
212 pp.

Contenido: INTRODUCCIÓN. PRIMERA PARTE. CAPÍTULO 1: QUINE Y LA OBSESIÓN INTENSIONAL. 1.1. El contexto de la crítica de las modalidades. 1.2. Entre intensionalidad y esencialidad. 1.3. Del esencialismo ingenuo al esencialismo advertido. 1.4. Cómo usar la distinción entre nombres y descripciones. CAPÍTULO 2: ALCANCE Y LÍMITES DE LA CRÍTICA DE QUINE. 2.1. Las fórmulas-Barcan: El "de re" y el "de dicto". 2.2. El argumento de Smullyan en defensa de las modalidades. SEGUNDA PARTE. CAPÍTULO 3. LA TEORÍA DE LOS MUNDOS POSIBLES. PARA UNA NUEVA FUNDAMENTACIÓN SEMÁNTICA DE LA LÓGICA MODAL. 3.1. Hintikka, Lewis, Kripke: tres concepciones de los mundos posibles. 3.2. Los modelos de Kripke. 3.3. Los contraejemplos a la fórmulas-Barcan. 3.4.. Fallas semánticas o problemas filosóficos. Exploraciones metasémánticas en los mundos posibles. CAPITULO 4: METÁFORAS Y MODELOS. LA METAFÍSICA DE LA LÓGICA MODAL. 4.1. Estructuras-modelo, trasidentificación y designadores rígidos. 4.2. Examen de la teoría causal de la referencia. 4.3. El test de Kripke y la vuelta al esencialismo. CAPÍTULO 5: A LA SOMBRA DE KANT.KRIPKE Y LO CONTINGENTE "A PRIORI": 5.1 De milagros filosóficos. El apareamiento de contingencia y aprioricidad. 5.2. ¿ Hay verdades contingentes cognoscibles a priopri?. 5.3. Kripke, Euler y los puentes de Königsberg. TERCERA PARTE. CAPÍTULO 6: SOBRE ESENCIAS, EPISTEMOLOGÍA Y REALISMO. 6.1. Integridad óntica y propiedades esenciales. 6.2 Bases objetivas de la relatividad de la identidad. CAPITULO 7: ESENCIALISMO. 7.1. Reconsideración del compromiso esencialista..7.2 Significatividad, "de re", "de dicto", oraciones esenciales.. CONCLUSIÓN. NOTAS. BIBLIOGRAFÍA.

González Carlomán, Antonio. *Lógica matemática para niños.*
Servicio de Publicaciones de la Universidad de Oviedo, Oviedo, 1991. 223 pp.

*Contenido.* 1. Introducción: Postura de Piaget; Postura de Wallon; Postura de Vigotsky; Observaciones. 2. Capítulo I: Lógica de proposiciones: Proposiciones simples; Proposiciones compuestas; Valoración de proposiciones; Universo del discurso; Tautologías (contradicciones); Teoremas; Lógica proposicional dirigida a los niños de los primeros cursos de básica. 3. Capítulo II: La lógica de cuantificadores o de predi-

cados: Cuantificadores; Lógica de cuantificadores dirigida a los niños de los primeros cursos de básica. 4. Capítulo III: Conjuntos: Definición y notaciones; Conjuntos deducidos de otros. 5. Capítulo IV: Algebra de Boole: axiomática de álgebra de Boole; Álgebra de Boole de los conjuntos de un universo U; Álgebra de Boole del conjunto de las proposiciones.

Gros, Begoña(coordinadora): *Diseños y programas educativos. Pautas pedagógicas para la elaboración de software.*
Ed. Ariel, Barcelona, 1997.
155 pp.

*Contenido:* Prólogo. Presentación: I. Un poco de historia. II. Objetivo de la Obra. III. Orientaciones para la lectura. 1. Las teorías sobre el diseño de software educativo: 1. Diseños y programas. 2. Tipos de productos. 3. La elaboración de software educativo. 3.1. El modelo sistemático. 3.2. Los modelos no lineales. 3.3. Los modelos hipertextuales. 4. Del conductismo al constructivismo. Referencias bibliográficas. Primera Parte: Aproximación conductista al diseño de software educativo. 1. Introducción. 2. Principios básicos del aprendizaje. 2.1. El condicionamiento operante. 2.2. Moldeamiento y generalización. 2.3. Reforzadores, refuerzo y contingencia. 2.4. Programas de refuerzo. 3. Principios básicos de la enseñanza. 4. La utilización de las teorías conductistas en el diseño de software educativo. Referencias bibliográficas. Segunda Parte: Aproximación cognitivista al diseño de software educativo. 3. La teoría de Robert Gagné. 1.Introducción. 2. Elementos de la teoría del aprendizaje. 2.1. Condiciones internas. 2.2. Condiciones externas. 2.3.Los resultados del aprendizaje. 3. Teoría de la instrucción. 3.1. Análisis de las tareas. 3.2. Organización de la instrucción para cada fase del aprendizaje. 4. La utilización de la teoría de Gagné en el diseño de software. Referencias bibliográficas. 4. La teoría de David Merril. 1. Introducción. 2. La teoría del Proceso Instructivo: Component Display Theory. 2.1. Clasificación de los resultados de la instrucción. 2.2.formas de presentación de la información. 3. La Teoría del diseño Instructivo: Component Design Theory. 3.1. Supuestos acerca del diseño instructivo. 3.2. La teoría de la transacción instructiva. 4.La utilización de la teoría de Merril en el diseño de software educativo. Referencias bibliográficas. Tercera parte: Aproximación constructivista al diseño de software educativo. 5. Las teorías constructivistas. 1. Introducción. 2. La teoría constructivista del aprendizaje. 2.1 Tendencias constructivistas: ¿ Son todos los constructivismos iguales? 2.2. Niveles de adquisición del conocimiento. 2.3. Entornos del aprendizaje. 3. Diseño instructivo. 3.1. Mayor énfasis en el aprendizaje y no en la instrucción. 3.2. Una propuesta diferente para el uso de la tecnología. 3.3. Propuesta de un di-

seño instructivo diferente. 4.La utilización de las teorías constructivistas en el diseño de software educativo. 4.1. Aplicación de la teoría de la flexibilidad cognitiva. Referencias bibliográficas. Cuarta parte: ¿Qué teoría para qué aprendizaje? 6. Cómo diseñar un programa educativo: un caso práctico. 1. Descripción de las características del programa. 2.La aplicación de las teorías conductistas. 2.1. Elección del formato del programa. 2.2. Estructura del programa. 2.3. Tipos de aprendizaje. 2.4. Tipos de estrategia de enseñanza. 2.5. Descripción del ejemplo. 3. La aplicación de las teorías cognitivas. 3.1.Pautas para la aplicación de la teoría de Gagné. 3.2. Pautas para la aplicación de la teoría de Merril. 3.3. Diseño del programa. 4. La aplicación de la teorías constructivistas. 4.1. Elección del formato del programa. 4.2. Tipos de aprendizaje. 4.3. Estrategias de enseñanza. 4.4. Descripción del ejemplo. 7. Recomendaciones para la utilización de las teorías del diseño de software educativo. 1. La importancia del diseño. 2. Una sola teoría no es suficiente. 2.1. Tipo de contenido. 2.2. Edad del usuario. 2.3. Contexto de uso.

Hausmann, Hanny. *El niño: El futuro ciudadano.*
Primera Edición. Monte Ávila Editores, Caracas, 1991. 353 pp.

*Contenido*. 1. Introducción. 2. Metas del programa de estudios sociales. 3. Planificación de los estudios sociales. 4. Enseñanza–Aprendizaje de los aspectos cognoscitivos de las ciencias sociales. 5. Enseñanza–Aprendizaje de las habilidades. 6. Enseñanza–Aprendizaje de la clarificación de valores. 7. Enseñanza–Aprendizaje del análisis y toma de decisiones. 8. Participación y acción social. 9. Enseñanza–Aprendizaje de los problemas contemporáneos. 10. Recursos, actividades y experiencias creativas para la Enseñanza–Aprendizaje. 11. Evaluación.

Johnson-Laird, P.N, J. Byrne, R.M y Evans, J. St. B. T: *Razonamiento y racionalidad. ¿Somos lógicos?*
Ed. Paidós, Barcelona, 1997.
182 pp.

*Contenido:* Prólogo. *1. Pensamiento y razonamiento proposicional. 2. La lógica proposicional.* 1. Proposiciones. 2. Conectores. 3. Tablas de verdad. 4. Validez de los argumentos. *3. Investigaciones psicológicas sobre razonamiento proposicional. 4. Razonamiento condicional.* 1. El condicional en el lenguaje natural. 2. Investigaciones psicológicas sobre el condicional. 2.1. Aceptación de las reglas de inferencias básicas. 2.2. "Si *p* entonces *q*" y " *p* sólo si *q*". 2.3. Tablas de verdad del condicional. 2.4. Influencia del contenido y del contexto. 3. Teorías sobre el condicional. 3.1. Teorías que proponen la existencia de reglas formales de inferencia. 3.2. Teoría de los modelos mentales.

3.2.1. Nivel computacional. 3.2.2. Nivel algorítmico. 3.2.3. Construcción de modelos para *modus ponens* y *modus tollens*. 3.2.4. Conjunto de modelos para condicionales. 3.2.5. Fenómenos que pueden ser explicados por la teoría. 5. La tarea de selección de Wason. 5.1. Descripción de la tarea. 5.2. Investigaciones que utilizan regla con contenido abstracto. 5.2.1. Variables que influyen en el rendimiento. 5.2.1.1. Variaciones en el enunciado de la regla y en las tarjetas utilizadas. 5.2.1.2. Efecto de las instrucciones. 5.2.1.3. Influencia del contexto. 5.2.1.4. Efecto de otras variables. 5.2.2. Investigaciones que utilizan reglas con contenido concreto. 5.2.2.1. Las reglas y su influencia en el rendimiento. 5.2.2.2. Las teorías explicativas. 5.2.2.2.1. Disponibilidad. 5.2.2.2.2. Esquemas de razonamiento pragmático. 5.2.2.2.3. teoría del contrato social. 5.2.2.2.4. teoría de la utilidad subjetiva. 5.2.2.2.5. Teoría de la detección de señales. 5.2.2.2.6. Teorías atencionales. 6. Razonamiento disyuntivo. 6.1 La disyunción en la lógica y en el lenguaje natural. 6.2. Investigaciones psicológicas sobre la disyunción. 6.2.1. Investigaciones sobre tablas de verdad. 6.2.2. Aceptación de las inferencias disyuntivas. 6.3.Teorías explicativas. 6.4. El problema THOG. 7. Lecturas complementarias. Deducción. 7.1. Implicaciones de la teoría de los modelos. 7.1.1. La adquisición de la competencia deductiva. 7.1.2. Los modelos mentales y otros modos de pensamiento. 7.1.3.Racionalidad versus relativismo. 7.2. Críticas a los modelos mentales. 7.2.1. Críticas a la teoría original. 7.2.2. Deficiencias de la teoría actual. Bibliografía. 8. Teorías del razonamiento humano: un panorama fragmentado. 8.1. La investigación en razonamiento deductivo. 8.1.1. La cuestión de la competencia. 8.1.2. La cuestión del sesgo. 8.1.3. La cuestión del contenido. 8.2. Explicaciones contemporáneas. 8.2.1 Explicaciones de la competencia. 8.2.2. Explicación del sesgo. 8.2.3. Explicaciones de los efectos del contenido y del contexto. 8.2.4. Conclusiones. 8.3. Cuestiones subyacentes. 8.3.1. Razonamiento versus no razonamiento. 8.3.2. La importancia de las representaciones mentales. 8.3.3. La naturaleza de los procesos de razonamiento. 8.4. Conclusiones y recomendaciones. 8.5. Bibliografía. 9. Epílogo. 9.1. La polémica actual entre teorías competitivas. 9.2. Los procedimientos de observación: tareas, evaluación, de las respuestas y teorías normativas. 9.3. El binomio competencia-ejecución y la discusión sobre racionalidad. 9.4. Diferencias individuales y niveles educativos. Bibliografía.

Lipman, M., Sharp, A.M., Oscanyan. *La filosofía en el aula.*
Ediciones de la Torre, Madrid, 1992. 379 pp.

*Contenido.* 1. Introducción. 2. Capítulo 1: Reconstruir los fundamentos. 3. Capítulo 2: La práctica filosófica y la reforma educativa. 4. Capítulo

3: La necesidad de una transformación educativa. 5. Capítulo 4: El pensamiento y el curriculum escolar. 6. Capítulo 5: La filosofía: La dimensión perdida de la educación. 7. Capítulo 6: Preparar al profesorado para enseñar a pensar. 8. Fines y métodos de filosofía para niños. 9. Capítulo 7: Algunas presuposiciones educativas de filosofía para niños. 10. Capítulo 8: El curriculum de filosofía para niños. 11. Capítulo 9: Metodología de la enseñanza: Consideraciones de valor y estándares de la práctica. 12. Capítulo 10: Dirigir una discusión filosófica. 13. La verdad, el bien y la belleza. 14. Capítulo 11:Animar a los niños a que sea lógicos. 15. Capítulo 12: ¿Se puede separar la educación moral de la investigación filosófica? 16. Capítulo 13. Educación para los valores cívicos. 17. Capítulo 14: Filosofía y creatividad. 18. Epílogo. 19. Capítulo 15: La filosofía de la infancia.

Marina, José Antonio. *El vuelo de la inteligencia.*
Primera Edición. Plaza & Janes Editores, Barcelona, 2000. 220 pp.

*Contenido.* 1. Aprender a aprender, por Margarita Revière. 2. La inteligencia resuelta. 3. La inteligencia y el lenguaje. 4. El bello discurrir de un sutil río. 5. ¿Y si el corazón se queda? 6. La inteligencia compartida. 7. El gran proyecto. 8. Antología de textos a modo de bibliografía: Tres rumbos del vuelo del águila; Primer rumbo, transfigurar el significado; Segundo rumbo, conocer; Tercer rumbo, transformar la realidad.

Revista: *Apuntes Filosóficos,* N° 15, 1999.
Consejo de Desarrollo Científico y Humanístico. Escuela de filosofía de la Universidad Central de Venezuela. Universidad Central de Venezuela.

*Contenido:* 1. Artículos: "El vicioso deseo del tirano platónico". W. Gil. "El poder en Aristóteles". A. Hermosa. "Filosofía y política en la defensa de la *naturalis contemplatio* en un aristotélico del renacimiento: Cesare Cremonini (1550- 1631)". G. Pagallo. "Kant y el humanismo". A. Renaut. "Observaciones y reflexiones en torno al tema de las relaciones entre creencia religiosa y racionalidad". C. Paván. "De la ciencia del hombre a la razón histórica; Filosofía y democracia ¿ cuál tiene la prioridad?". O. Astorga y M. Cisneros. "Leopoldo Zea y la filosofía de la historia". H. Jaimes. "Ética y estética" E. Heymann. Documentos: F. Bacon: Prometeo o la situación del hombre. Reseñas: Después del fin del arte (Guadalupe Llanes).

Revista: *Araucaria (Revista Iberoamericana de Filosofía, Política y Humanidades).*
Niño y Dávila Editores y Universidad de Sevilla, Año 1, N°2. Segundo semestre de 1999.
206 pp.

*Contenido:* **Las ideas: su política y su historia: "**La imaginación y la estructura del pensamiento político de Hobbes". O. Astorga. "¿ Es posible una ciudadanía global?. A. Colomos. "Populismo y nacionalismo" G. Hermet. **Monográfico: El nuevo presidencialismo en Iberoamérica.** "Multipartidismo, federalismo robusto y presidencialismo en Brasil". S. Mainwaring. "Crítica al presidencialismo en América Latina". E. Bernales. **Perfiles/Semblanzas:** "Los peligros de la distracción (A propósito de Jorge Millas a diez años de su muerte". H. Giannini. "Vestigios de doscientos años: México, historia política y biografía". M. Sáez. "Filosofía y democracia: una relación más complicada". R. Rodríguez. **Documentos:** "La concepción de libertad - poder de Friedrich von Hayec". J. Millas.

Revista: *Filosofía (Revista del postgrado de filosofía de la Universidad de Los Andes)*
Consejo de publicaciones-ULA, N°11, Tomo 1, 1999.
523 pp.

*Contenido:* **La responsabilidad del filósofo en tiempos de crisis (Actas del IV Congreso Nacional de Filosofía):** "Tiempo de historia y de filosofía" Jesús Rondón Nucete. "La responsabilidad del filósofo en tiempos de crisis". Ernesto Mayz Vallenilla, Benjamín Sánchez y Eduardo Pianza. **Metafísica:** "Ser y sustancia en la metafísica de Aristóteles". B. Bordalejo. "El problema del sustrato en el capítulo III del libro VIII de la metafísica de Aristóteles". M. Llorens. "Gilson lectorde Santo Tomás: Apuntes metafísicos". C. Paván. "Libertad y autonomía del sujeto (subjetividad en el ocaso de la conciencia)". R. Hurtado. **Filosofía de la historia:** "Categorías teóricas del análisis histórico". A. Orcajo. "Filosofía de la historia en Tucídides: su concepción del sentido del devenir humano". G.R. Quintero L. "Tradición y lenguaje: La historia como construcción de la humanidad". T. Bianculli. "Tiempos de crisis, tiempo de oportunidad". A. Gandara. **Filosofía del lenguaje** "Intención y convención en los actos de habla: una contribución de Strawson a la teoría de los actos de habla". Nancy Núñez. "Credo semántico de un incomensurabilista coherente". J. Nicolic. "Quine: Ontología de lo necesario". T. Olmos. "Modalidad y esencialismo: sobre el alcance de la crítica quiniana". V.P. Lo Mónaco. "El lenguaje: nuestro límite con Dios (acercamiento a la filosofía del lenguaje en Jonuel Brigue". A. Rodríguez Silva. "Cambios de teoría en la lingüística del

siglo XX a la luz de los planteamientos de Karl Popper. **Epistemología de las ciencias humanas:** "León Wygotsky: significación y sentido personal en la actividad humana". L. Alonso. "Racionalidad y discurso de las necesidades sociales". A. Márquez. "Anomalismo y racionalización de las acciones humanas". V. Rodríguez. "Filosofía y educación: reexamen de su relación histórica y conceptual". L. Molina. **Antropología y Filosofía política de Hobbes y Kant:** "El concepto hobbesiano de placer en el Leviatán". E. Gonzáles. "Del Leviatán a la nueva síntesis: Aspectos de la relación entre T. Hobbes y la socio-biología humana". I. Argibay. "Reexamen de la filosofía política de Kant ( desde Hobbes)". O. Astorga. "La antropología en la filosofía trascendental kantiana". F. Zambrano. **Proceso de la 'Razón de Estado' en la soceidad contemporánea:** "¿Obedecen tanto lo económico como lo político a una misma racionalidad?". M. Laporte. "La responsabilidad del evolucionismo social en la crisis contemporánea". M. del Pilar Quitero. "Poder y comunicación". R. Pernía. "Antecedentes filosófico–políticos del renacimiento islámico actual". L. Vivencio Saavedra. **Ética aplicada: filosofía jurídica y derechos humanos:** "Fuentes axiológicas del discurso político en la aplicación del uso alternativo del derecho en el área penal". J.L Rosell Senhenn. "Crisis, fractura cultural y filosofía: filosofía jurídica". A.J. Bozo de Carmona. "Naturaleza humana, derechos humanos y razón práctica jurídica". R. Carrión. "De la Ética a la Ética biomédica". B. Bernard. "En busca de una ética para la infancia". E. Urdaneta. "Hacia una antropología para las mujeres". E. Aponte. "Filosofía del derecho y bioética". D. Labarca. **Pensamiento filosófico latinoamericano:** "La axiómata Caracensia". A. Muñoz. "América, la utopía de Simón Rodríguez". C. Jorge. "Andrés Bello y los usos del filosofar". J. Sasso. "El pórtico de los libertadores. Un acercamiento a la influencia de la filosofía estoica en el pensamiento educativo del Libertador". M. Nava. **El debate "Modernidad y Postmodernidad":** "El poder de los filósofos". H. Calello. "Psicoanálisis y crisis: un recorrido en busca del sujeto perdido". S. Strozzi. "La situación epistémica de la ficción". R. Guzmán. "La ilusión de la antropología postmoderna". R. Ma. Di Falco. **Conferencia:** "Hacia una macroética de la corresponsabilidad". Dr. Karl Otto Apel.

Revista: *Logoi*
Centro de Estudios filosóficos UCAB, N° #, 2000.
224 pp.

*Contenido:* "La paz de un guerrero del pensamiento". J. Dávila. "Foucalt y la función intelectual: un janseísmo político". F. Gros. "De Nietzche a Foucalt, un peligrosos tal vez". J. Jara. "Apuntes para una crítica a la teoría liberal de la democracia: una visión desde América Latina". C. Kohn. "Reconstrucción del campo intelectual de las teorías sociológi-

cas del siglo XIX". M. Mujica Ricardo. "La esperanza social post-metafísica en Richard Rorty". O. Reyes. "Lo dionisíaco y lo apolíneo como elementos configuradores de lo humano (una lectura de "El nacimiento de la tragedia" de Nietzche)". C. Rivas. "Hermenéutica, postmodernidad, violencia". M. Desiato. "Un camino para filosofar en psicología: contribuciones de la antropología filosófica a la metapsicología". C. Rivas. "A propósito de imposturas intelectuales: de Sokal y Bricmont". T. Hannot. *Reseñas:* "La tercera vía. La renovación de la socialdemocracia". "Ciudadanía y clase social". "El déficit social neoliberal". "El paroxista indiferente. Conversaciones con Philippe Petit". "*Performing Psychology: A postmordern culture of the mind*". Actualidad filosófica. Índice acumulado.

Revista*: Tharsis: Reflexiones sobre la educación, ética y sociedad* N° 5/6, 1999.
Consejo de Desarrollo Científico y Humanístico de la Universidad Central de Venezuela, Caracas, 1999. 170 pp.

*Contenido:* "La Universidad y el proyecto republicano venezolano del siglo XIX". A. Navas Blanco. "La crisis como trama: notas sobre la universidad contemporánea". M. Téllez. "El reto del desarrollo educativo del personal académico de las universidad nacionales". F. Nieves. "Las políticas públicas en la educación superior venezolana: o la comedia de las equivocaciones". J.M. Cortázar. "Gerencia del conocimiento, ética y descentralización". G. Portillo. "Moral y derecho". E. Vásquez. "Atenas y Jerusalén (¿Qué significa ser humano?)". E. Gómez. "La crítica de Rosseau a la sociedad contemporánea". A.H. Andújar. "El crepúsculo de la política". L. Salamanca. "Ciudadanía y crisis de la política". R. Orta. "El laberinto circular: Borges Viquiano". J.R. Herrera. "Latinoamérica y complejo de inferioridad". M. Padrón. "Los intelectuales venezolanos y el despertar de la conciencia nacional entre 1928 y 1935". L.R Dávila.

Savater, Fernando: *Las preguntas de la vida.*
Ed. Ariel, Barcelona, 1999.
286 pp.

*Contenido:* Introducción: El por qué de la filosofía. 1. La muerte para empezar. 2. Las verdades de la razón. 3. Yo adentro, yo afuera. 4. El animal simbólico. 5. El universo y sus alrededores. 6. La libertad en acción. 7. Artificiales por naturaleza. 8. Vivir juntos. 9. El escalofrío de la belleza. 10. Perdidos en el tiempo. Epílogo: La vida sin por qué. Despedida. Principales estrellas invitadas.