

La spiegazione dei processi cognitivi*

1. IL PROBLEMA

Che tipo di spiegazione sistematica dovremmo e possiamo trovare nella scienza cognitiva per quanto riguarda la percezione, la comprensione linguistica, l'azione razionale e le altre forme di cognizione? A grandi linee penso che la risposta sia ragionevolmente chiara: stiamo cercando spiegazioni causali e il nostro oggetto sono *certe funzioni* di un organo biologico, il cervello umano e animale.

Come per ogni altra scienza naturale vi sono certe assunzioni che dobbiamo fare e certe condizioni che le nostre spiegazioni devono soddisfare. Specificamente, dobbiamo supporre che esista una realtà totalmente indipendente dalle rappresentazioni che ne diamo (in un'era intellettualmente più sana non sarebbe necessario dirlo) e dobbiamo supporre che gli elementi di quella realtà a cui ci riferiamo nelle nostre spiegazioni funzionino genuinamente in modo causale.

Non tutte le funzioni del cervello sono rilevanti per i processi cognitivi, sicché dobbiamo essere cauti nel restringere la gamma delle funzioni del cervello di cui discutiamo. La scienza cognitiva tratta del funzionamento *cognitivo* del cervello e delle sue relazioni con il resto dell'organismo e con il resto del mondo nella maniera in cui la scienza della nutrizione tratta del funzionamento digestivo e delle sue relazioni con il resto dell'organismo e con il resto del mondo. Come gli altri organi, e invero come gli altri sistemi fisici, il cervello presenta livelli differenti di descrizione e la scienza cognitiva ha a che fare propriamente con ogni livello di descrizione del cervello che sia rilevante per la spiegazione causale dei processi cognitivi. Questi possono andare dai processi coscienti di decisione, al livello più alto, fino alla struttura molecolare dei neurotrasmettitori, al livello più basso.

Tipicamente, i livelli superiori saranno proprietà causalmente emergenti del comportamento e dell'organizzazione degli elementi del cervello ai livelli inferiori. Si consideri un ovvio esempio di senso comune di una spiegazione

* Traduzione dall'inglese di Baldassare Pastore.

relativa a uno di questi livelli superiori. Se io cercassi di spiegare il mio comportamento di guida in Gran Bretagna dicendo che sto seguendo la regola “guida a sinistra”, fornirei una vera e propria spiegazione causale riferendomi a un processo mentale. L’operare della regola è di per sé causato da eventi neuronali di livello inferiore nel cervello ed è realizzato nel cervello a un livello più alto di quello dei singoli neuroni. In un senso spero non misterioso di “proprietà emergente”, l’operare della regola nel produrre il mio comportamento è una proprietà causalmente emergente del sistema cerebrale. Un’altra maniera di porre tale argomento equivale a dire: possiamo fornire vere e proprie spiegazioni causali che non stanno al livello inferiore, al livello dei neuroni ecc., poiché anche i livelli superiori di spiegazione sono livelli reali. Parlarne non è solo un modo di dire o una metafora. Perché sia un livello reale, un supposto livello causale deve essere propriamente collegato ai livelli più fondamentali, per esempio ponendosi come proprietà causalmente emergente di tali livelli. Chiamiamo questo vincolo, ossia il fatto che nello spiegare i processi cognitivi dobbiamo riferirci a caratteristiche reali del mondo reale che funzionano causalmente, *vincolo di realtà causale*.

Così, per riassumere questi vincoli, cerchiamo *spiegazioni causali* del cervello *funzionanti* a differenti *livelli di descrizione*. Ci concediamo una completa libertà nel parlare di differenti livelli di descrizione, ma tale libertà è vincolata dall’esigenza che i livelli siano causalmente reali.

Ciò che qui intendo sostenere è che alcuni modelli esplicativi nella scienza cognitiva, sebbene naturalmente non tutti, non riescono a soddisfare il vincolo di realtà causale. Intendo inoltre proporre alcune revisioni che renderebbero possibili le spiegazioni al fine di soddisfare quel vincolo.

2. LA VERSIONE DEL MODELLO DELL’ELABORAZIONE DI INFORMAZIONI FORNITA DA MARR

Ludwig, il mio cane, è molto bravo nell’afferrare al volo palle da tennis. Per esempio, se fai rimbalzare una palla da tennis contro un muro, egli è solitamente capace di fare un balzo e porre la sua bocca nel punto preciso in cui la palla arriva stringendola tra i denti. Non gli succede sempre, ma è abbastanza bravo in ciò. Come riesce a farlo?

Secondo i modelli esplicativi correnti nella scienza cognitiva, Ludwig esegue un compito di elaborazione di informazioni enormemente complesso. Raccoglie informazioni nella forma bidimensionale sulla sua retina, le elabora attraverso il sistema visivo fino a produrre una rappresentazione tridimensionale del mondo esterno e immette quella rappresentazione nel sistema motorio in uscita. Le computazioni che egli esegue, persino per l’unità di movimento in uscita, non sono cose da nulla. Qui abbiamo un candidato per la prima formulazione di uno degli algoritmi. Ludwig inconsciamente sta

seguendo la regola: salta in modo tale che il piano dell'angolo di riflessione della palla sia esattamente eguale al piano dell'angolo di incidenza dell'impatto, e metti la tua bocca nel punto in cui la palla è in un arco parabolico, l'uniformità della cui traiettoria e della cui velocità è una funzione dei tempi di velocità d'impatto, del coefficiente di elasticità della palla da tennis, meno una certa perdita dovuta all'attrito dell'aria. Ossia, in base al modello computazionale standard di cognizione, Ludwig computa inconsciamente un gran numero di funzioni facendo in maniera inconscia un sacco di matematica.

All'apparenza, la spiegazione del suo comportamento è proprio quella della persona che segue la regola "guida a sinistra" tranne per il fatto che non c'è modo, neanche in linea di principio, che egli possa diventare consapevole in maniera cosciente dell'operare della regola. Le regole non solo non sono di fatto presenti alla coscienza, ma non sono neppure il genere di regole che egli può diventare consapevole di seguire. Esse sono ciò che ho chiamato regole "profonde inconse"¹.

Non sono mai stato completamente soddisfatto di questo tipo di spiegazione. Il problema non è soltanto che esso attribuisce una terribile quantità di conoscenza matematica inconscia al cervello canino di Ludwig, ma la cosa più importante è che tralascia l'elemento cruciale consistente nel fatto che Ludwig è un agente razionale cosciente che prova a fare qualcosa. Un siffatto modello esplicativo sembra più appropriato per chi sta costruendo una macchina, un robot canino, che afferrerebbe al volo palle da tennis. Ritengo infatti che il fascino intuitivo dell'approccio risieda nel fatto che esso predirebbe il comportamento di Ludwig e ciò è il tipo di informazione che vorremmo inserire in un robot se noi lo stessimo costruendo per simulare il suo comportamento.

Allora sondiamo un po' più a fondo le assunzioni che stanno dietro questo approccio.

La formulazione classica di questa versione del paradigma esplicativo della scienza cognitiva è dovuta a David Marr², ma vi sono prospettive equivalenti in altri autori. In base a questo paradigma la scienza cognitiva è un tipo speciale di scienza dell'elaborazione di informazioni. Siamo interessati a come il cervello e gli altri sistemi funzionalmente equivalenti, come certi tipi di computer, elaborano informazioni. Vi sono tre livelli di spiegazione. Il più alto è il livello computazionale, e questo Marr lo definisce in termini di vincoli di informazione disponibili per tracciare un percorso da una informazione in entrata a una in uscita. Nel caso di Ludwig il compito computazionale per il suo cervello consiste nel ricevere informazioni circa un assetto visivo bidi-

1. J. R. Searle, *La riscoperta della mente*, trad. it. di S. Ravaioli, Bollati Boringhieri, Torino 1994, cap. 7.

2. D. Marr, *Vision*, Freeman and Co., San Francisco (CA) 1982.

mensionale e produrre rappresentazioni delle contrazioni muscolari che porterebbero la sua bocca e la palla da tennis nello stesso posto allo stesso momento.

Intuitivamente penso che l'idea di Marr del livello computazionale sia chiara. Se si istruisce un programmatore di computer per progettare un programma, la prima cosa che gli si dice è quale attività si vuole che il programma svolga. E il resoconto di quell'attività è un resoconto del compito computazionale da eseguire al livello computazionale.

Come avviene questo? Bene, ciò conduce al secondo livello, che Marr definisce livello algoritmico. L'idea è la seguente. Ogni compito computazionale può essere eseguito in modi differenti. L'idea intuitiva è che il livello algoritmico determina il modo in cui il compito computazionale è eseguito da uno specifico algoritmo. In un computer penseremmo al livello algoritmico come quello riguardante il programma.

Un aspetto problematico delle versioni della scienza cognitiva riguardo a questo livello è la dottrina della scomposizione ricorsiva. I livelli complessi si scompongono in livelli più semplici finché non viene raggiunto il livello più basso e a quel livello è tutta una questione di zero e uno, o di altri simboli binari. Cioè, non vi è realmente un livello algoritmico intermedio singolo, ma piuttosto una serie di livelli a "scatola cinese" che terminano nei processori primitivi, e questi sono simboli binari. E il livello inferiore è l'unico a essere reale. Tutti gli altri sono riducibili a esso. Ma anche *questo* non possiede alcuna realtà fisica. È implementato nella fisica, come vedremo, però *il livello algoritmico non fa alcun riferimento ai processi fisici*.

Un tempo pensavo che le computazioni che attribuiamo a Ludwig potessero essere, in base a questo modello, il livello causalmente reale, ma non è così. Tutto ciò che Ludwig fa realmente è manipolare gli zero e gli uno. Tutto il resto è mera apparenza. Ogni affermazione computazionale che noi facciamo su Ludwig si riduce all'affermazione che egli sta manipolando gli zero e gli uno.

Per Marr il livello più basso è il livello di implementazione, relativo a come l'algoritmo è effettivamente implementato in uno specifico hardware. Lo stesso programma, lo stesso algoritmo, può essere implementato in una gamma indefinita di differenti hardware ed è ben possibile, per esempio, che un programma implementato nel cervello di Ludwig possa essere anche implementato su un computer commerciale.

Così, sulla base del modello tripartito di Marr si ottiene il seguente quadro. La scienza cognitiva è essenzialmente la scienza dell'elaborazione di informazioni in un senso molto particolare di tale nozione, e riguarda primariamente la spiegazione del livello superiore attraverso il livello algoritmico. Ciò che conta per la spiegazione della scienza cognitiva è il livello intermedio. Perché? Perché dovremmo spiegare i cervelli al livello intermedio e non al

livello dell'hardware? La risposta è data dalla mia iniziale caratterizzazione del cervello come sistema *funzionale*. Laddove abbiamo a che fare con altri sistemi funzionali, come le automobili, i termostati, gli orologi e i carburatori, siamo interessati al modo in cui la funzione è eseguita al livello della funzione, non al livello della microstruttura. Perciò nello spiegare il motore di un'automobile parliamo di pistoni e cilindri e non di particelle subatomiche delle quali il motore è composto; perché, parlando in maniera grossolana, ognuna delle vecchie particelle subatomiche andrà bene purché implementi i pistoni e i cilindri. Nel caso di Ludwig siamo interessati alla regola inconscia che egli effettivamente sta seguendo e non all'implementazione neuronale del comportamento consistente nel seguire la regola. E la regola che egli sta effettivamente seguendo può essere interamente espressa in termini di zero e uno, perché questo è quanto sta realmente accadendo. Così, in base a questa concezione, la mia precedente caratterizzazione della scienza cognitiva come scienza della funzione celebrale a un certo livello o a certi livelli di descrizione era fuorviante. La scienza cognitiva è una scienza di elaborazione di informazioni, che capita sì implementi nel cervello ma che potrebbe egualmente ben implementarsi in una gamma indefinita di altri hardware. La scienza cognitiva spiega il livello superiore nei termini del livello intermedio ma non è realmente interessata al livello inferiore eccetto per il fatto che questo implementa il livello intermedio.

Un problema dell'analisi tripartita della funzione cognitiva fornita da Marr riguarda proprio il fatto che ogni sistema ammetterà questo stile di analisi. E il punto è che non solo gli orologi, i carburatori e i termostati ammettono l'analisi a tre livelli (ciò è ben accetto da parte dei seguaci del modello classico come dimostrazione che i processi cognitivi ammettono un'analisi funzionale simile a quella degli orologi ecc.). Il problema è che ogni sistema di una qualche complessità ammette un'analisi dell'elaborazione di informazioni.

Si consideri un sasso che cade dalla cima di un dirupo. Il "sistema", se posso così descriverlo, presenta tre livelli di analisi. Il compito computazionale per il sasso consiste nel calcolare un modo per arrivare al suolo in un certo lasso di tempo. Deve computare la funzione $S = 1/2 gt^2$. Al livello intermedio c'è un algoritmo che esegue il compito. L'algoritmo istruisce il sistema riguardo ai passi da compiere per mettere in corrispondenza tempo e spazio nel modo corretto. E qui vi è la familiare implementazione hardware in termini di massa del sasso, terra e aria frapposta. Allora perché il sasso che cade non è un sistema di elaborazione di informazioni? Se invece il sasso lo è, allora lo è ogni cosa.

Si tratta di una questione cruciale per la scienza cognitiva e svariati autori ne hanno dato una risposta. Stando a quanto dicono, abbiamo bisogno di distinguere tra un sistema che può essere descritto per mezzo di una compu-

tazione e uno che effettivamente esegue la computazione. Il sistema sopra citato *può essere descritto* per mezzo di una funzione calcolabile, ma non *esegue* quella computazione perché (a) non vi sono rappresentazioni per la computazione da effettuare e (b) *a fortiori* non vi è alcuna informazione codificata nelle rappresentazioni. Un'autentica scienza dei processi cognitivi, una scienza dell'elaborazione di informazioni richiede computazioni eseguite su simboli o su altri elementi sintattici, e queste sono le rappresentazioni che codificano l'informazione che è elaborata dall'algoritmo. Queste condizioni non sono soddisfatte da un sasso che cade anche se il sasso può essere descritto computazionalmente.

Se abbiamo l'intenzione di fornire una risposta precisa, abbiamo bisogno di una definizione o di un resoconto soddisfacente di "informazione", "rappresentazione", "simbolo" e "sintassi", non di riferirci a "computazione" e "algoritmo". E questi resoconti devono metterci in grado di spiegare come l'informazione, la rappresentazione ecc. entrino nel sistema in modo da soddisfare il vincolo di realtà causale. Il resoconto dovrà mostrare, in primo luogo, come l'informazione entri nel sistema in qualche forma intrinseca, e poi conservi il suo carattere di informazione per tutta l'elaborazione. Inoltre il resoconto dovrebbe mostrare come il livello reale dell'elaborazione di informazioni sia una proprietà emergente dei più fondamentali microlivelli. Per riferirci a casi specifici, non sarà sufficiente dire, come ha fatto Marr, che c'è un assetto visivo bidimensionale sulla retina come input per il sistema; dobbiamo invece dire quale fatto relativo a un tale assetto visivo lo rende un'informazione e quale sia esattamente il contenuto dell'informazione.

Ho dato uno sguardo a moltissima letteratura in argomento e non riesco a trovare una definizione soddisfacente di rappresentazione o informazione o delle altre nozioni che risolva i nostri problemi. A loro merito, Palmer e Kimchi³ ammettono di non avere la più pallida idea di ciò che possa essere, secondo loro, informazione. Intendo esaminare la nozione di informazione un po' più dettagliatamente. La domanda fondamentale di questo scritto è: possiamo dare un senso empirico ai concetti fondamentali del modello dell'elaborazione di informazioni che renderebbe la versione dell'elaborazione di informazioni propria della scienza cognitiva una valida disciplina empirica?

3. SEGUIRE UNA REGOLA

Se vogliamo essere chiari circa l'affermazione che l'agente cognitivo sta seguendo regole inconsce, dobbiamo inizialmente comprendere cosa sia

3. Cfr. S. E. Palmer, *Vision Science: Photons to Phenomenology*, MIT Press, Cambridge (MA) - London 1999.

implicato nel comportamento consistente nel seguire regole. Si consideri un caso in cui sembra chiaro e certo che un agente sta seguendo una regola. Quando guido in Inghilterra, seguo la regola: guida sul lato sinistro della strada. E se sto in Inghilterra per un bel po' di tempo, constato che mi abito a guidare a sinistra sicché non devo pensare in maniera cosciente alla regola. Sembra naturale dire che sto ancora seguendo la regola anche quando non ci penso.

Una siffatta spiegazione soddisfa il vincolo di realtà causale. Quando dico che sto seguendo una regola sto dicendo che c'è un contenuto intenzionale a me intrinseco, il contenuto semantico della regola, che sta funzionando causalmente al fine di produrre il mio comportamento. Tale contenuto intenzionale è a un livello emergente dell'elaborazione cerebrale. La regola ha la direzione di adattamento mondo-a-regola e la direzione di causazione regola-a-mondo.

Voglio esplorare alcune delle caratteristiche di questo tipo di spiegazione per vedere se possono essere conservate nella scienza cognitiva dell'elaborazione di informazioni alla Marr. Farò semplicemente una lista di quelle che a me sembrano alcune caratteristiche importanti del comportamento consistente nel seguire regole.

1. La più importante singola caratteristica è quella non appena menzionata. Il contenuto intenzionale della regola deve funzionare causalmente nella produzione del comportamento in questione. Per far ciò deve essere a un livello emergente del funzionamento cerebrale. Questo chiarisce come le spiegazioni mediante regole nella vita reale soddisfino il vincolo di realtà causale. Anche ogni spiegazione in base al seguire regole nella scienza cognitiva deve soddisfare quel vincolo.

2. Il seguire regole ha carattere normativo dal punto di vista dell'agente. Il contenuto della regola determina per l'agente ciò che vale come giusto o sbagliato, in quanto riesce o meno.

3. La caratteristica successiva è una conseguenza della prima. La regola deve avere una certa forma aspettuale, ciò che Frege chiama il "modo di presentazione". Questo avviene perché regole estensionalmente equivalenti possono differire nella loro forza esplicativa. Posso seguire una regola e non un'altra, anche se il comportamento osservabile è lo stesso in entrambi i casi. Per questa ragione, poiché le spiegazioni mediante regole devono far menzione delle specifiche forme aspettuali, le spiegazioni mediante regole sono intensionali-con-la-s. Per esempio, la regola "su strade a due corsie guida a sinistra" è estensionalmente equivalente alla regola "guida in modo tale che il volante sia più vicino alla linea centrale della strada", data la struttura delle automobili inglesi che guido. Ma in Gran Bretagna seguo la prima regola e non la seconda, anche se ciascuna predirebbe altrettanto bene il mio comportamento.

4. Nel normale comportamento diretto da regole, le regole o sono coscienti o sono accessibili alla coscienza. Anche quando sto seguendo irriflessivamente la regola, potrei ancora pensarci. Non sempre sono cosciente della regola, ma posso facilmente diventarlo. Anche se la regola è così radicata nel mio inconscio da non riuscire a pensarci, deve risultare ancora possibile che sia quel tipo di cosa che può *essere* cosciente.

5. L'accessibilità alla coscienza implica una quinta esigenza. I termini nei quali la regola è espressa devono essere termini che si trovano nel repertorio cognitivo dell'agente in questione. È una caratteristica generale delle spiegazioni intenzionalistiche, di cui le spiegazioni mediante regole sono un caso speciale, che l'apparato a cui si fa ricorso per mezzo della regola deve essere quello di cui l'agente è in possesso. Se volessi spiegare perché Hitler invase la Russia, dovrei usare termini che fanno parte del repertorio concettuale di Hitler. Se postulassi alcune formule matematiche di cui Hitler non ha mai sentito parlare, che non ha potuto padroneggiare e di cui non ha potuto essere consapevole, allora la spiegazione non potrebbe essere una spiegazione intenzionalistica. È una peculiarità dei processi cognitivi, spesso sottolineata da chi discute dei caratteri specifici della spiegazione storica, che le spiegazioni che si rivolgono agli stati e ai processi cognitivi devono impiegare concetti disponibili all'agente.

6. La caratteristica che segue viene raramente sottolineata: seguire regole è normalmente una forma di comportamento volontario. Dipende da me seguire la regola o violarla. La regola infatti funziona causalmente, ma la regola in quanto causa, persino la regola insieme al desiderio di seguirla, non fornisce condizioni causalmente sufficienti.

Questo è tipico delle spiegazioni razionali del comportamento. Spesso si dice che le azioni sono causate da credenze e desideri. Ma se la consideriamo come un'affermazione circa le condizioni causalmente sufficienti, risulta falsa. Un test della razionalità del comportamento consiste nel fatto che c'è un divario tra i contenuti intenzionali (credenze, desideri, consapevolezza delle regole ecc.) e l'azione effettiva. Si deve comunque cambiare idea e fare la cosa che si è deciso di fare, anche nei casi in cui la regola esige che si faccia. Chiamerò questo divario tra la regola, e gli altri fenomeni intenzionali che sono le cause, e l'azione, che è il loro effetto, il "divario che caratterizza l'azione volontaria" o semplicemente "il divario".

7. Una caratteristica, collegata al divario, consiste nel fatto che le regole sono sempre soggette a interpretazione e all'interferenza di altre considerazioni razionali. Così, per esempio, non seguo la regola "guida sul lato sinistro della strada" ciecamente. Se c'è una buca, o un'automobile che blocca la strada, la eviterò. Tali regole sono in questo senso regole *ceteris paribus*.

8. L'ultima caratteristica consiste nel fatto che la regola deve operare in tempo reale. Per l'effettivo comportamento diretto da regole, la spiegazione

mediante regole richiede che il tempo dell'applicazione della regola e il tempo del funzionamento causale siano co-estensivi.

Per riepilogare, allora, abbiamo otto caratteristiche delle spiegazioni intenzionalistiche mediante regole. Primo, il contenuto intenzionale della regola deve funzionare causalmente; secondo, la regola stabilisce un modello normativo per l'agente. Terzo, le regole hanno una forma aspettuale e così le spiegazioni mediante regole sono intensionali-con-la-s. Quarto, la regola deve essere cosciente o accessibile alla coscienza. Quinto, le regole devono avere contenuti semantici propri del repertorio cognitivo dell'agente. Sesto, il comportamento diretto da regole è volontario, e perciò, a causa del divario che caratterizza il comportamento volontario, la spiegazione mediante regole non fornisce condizioni causalmente sufficienti. Settimo, le regole sono soggette a interpretazione e all'interferenza di altre considerazioni. Infine, la regola deve operare in tempo reale.

Confrontiamo tutto ciò con le forme cognitiviste di spiegazione alla Marr. In tali spiegazioni sono presenti in maniera inequivocabile solo le caratteristiche 1 e 3. Ora, un problema connesso con il vincolo di realtà causale riguardo le spiegazioni della scienza cognitiva consiste nel fatto che non è chiaro come si possano avere quelle due caratteristiche senza nessuna delle altre sei. Non è casuale che queste caratteristiche si tengano insieme, poiché le spiegazioni in base al seguire regole sono tipiche delle spiegazioni intenzionalistiche del comportamento razionale. Come può accadere in senso letterale che Ludwig stia seguendo una regola avente un particolare contenuto semantico, se tale regola non è per lui normativa, non è accessibile alla sua coscienza neanche in linea di principio, ha concetti totalmente al di fuori del suo repertorio, non è applicabile volontariamente, non è soggetta a interpretazione e sembra operare immediatamente piuttosto che in tempo reale?

4. ALCUNE DISTINZIONI PRELIMINARI

In questa sezione voglio rammentarvi alcune distinzioni fondamentali. Innanzitutto abbiamo bisogno di richiamare alla memoria la nota distinzione tra comportamento *diretto o guidato da regole*, da un lato, e comportamento *descritto da regole*, dall'altro. Quando seguo una regola, come, in Inghilterra, la regola del codice stradale "guida sul lato sinistro" l'effettivo contenuto semantico della regola gioca un ruolo causale nel mio comportamento. La regola fa di più che predire il mio comportamento; essa piuttosto è parte della causa del mio comportamento. Da questo punto di vista differisce dalle leggi di natura che descrivono il mio comportamento, comprese le sue cause, ma non causano il comportamento che descrivono. La distinzione tra guidato da regole e descritto da regole può essere generalizzata come una

distinzione tra *guidato dall'intenzionalità e descritto dall'intenzionalità*. Tutte le descrizioni hanno intenzionalità, ma la peculiarità delle spiegazioni intenzionalistiche dei processi cognitivi umani consiste nel fatto che il contenuto intenzionale della spiegazione funziona casualmente nella produzione dell'*explanandum*. Se dico "Sally bevve perché aveva sete", la sete funziona causalmente nella produzione del comportamento. *È importante tenere a mente questa distinzione perché se una scienza cognitiva dell'elaborazione di informazioni deve soddisfare il vincolo, l'intenzionalità dell'informazione non deve solamente descrivere ma deve funzionare causalmente nella produzione dei processi cognitivi che l'elaborazione di informazioni spiega. Altrimenti non c'è alcuna spiegazione causale.* Per soddisfare il vincolo di realtà causale, il livello algoritmico deve funzionare causalmente.

Credo che i resoconti standard della scienza cognitiva riconoscano questo punto quando distinguono l'essere descrivibile per mezzo di una funzione e il computare effettivamente una funzione. Questo è un caso speciale della distinzione generale tra descritto da regole e guidato da regole.

La seconda importante distinzione è tra le caratteristiche della realtà relative all'osservatore e quelle indipendenti dall'osservatore. Per la nostra visione scientifica globale del mondo risulta fondamentale la distinzione tra quelle caratteristiche che esistono indipendentemente da un osservatore, un progettista o un altro agente intenzionalistico e quelle che dipendono dagli osservatori, dagli utilizzatori ecc. Spesso lo stesso oggetto avrà entrambi i tipi di caratteristiche. Gli oggetti nella mia tasca hanno alcune caratteristiche indipendenti dall'osservatore come una certa massa e una certa composizione chimica, ma hanno anche caratteristiche relative all'osservatore: per esempio, uno è un biglietto da dieci sterline inglesi e un altro è un coltello dell'esercito svizzero. Intendo descrivere questa distinzione come quella tra caratteristiche del mondo che sono relative all'osservatore (o all'intenzionalità), e caratteristiche che sono indipendenti dall'osservatore (o dall'intenzionalità). Soldi, proprietà, matrimonio, governo e una corretta pronuncia inglese, così come i coltelli, le vasche da bagno e le automobili sono relativi all'osservatore; forza, massa e attrazione gravitazionale sono indipendenti dall'osservatore.

"Relativo all'osservatore" non significa arbitrario o irrealistico. Il fatto che qualcosa sia un coltello o una sedia o una bella giornata per un picnic è relativo all'osservatore ma non è arbitrario. Non si può usare qualsiasi cosa come coltello o sedia o come bella giornata per un picnic. Il punto che riguarda la relatività rispetto all'osservatore è che le caratteristiche relative all'osservatore, in tali descrizioni, esistono solo relativamente a osservatori umani. Il fatto che questo oggetto nella mia mano abbia una certa massa non è relativo all'osservatore ma indipendente dall'osservatore. Che lo stesso oggetto sia un coltello è relativo al fatto che agenti umani lo hanno progettato, venduto, usato

ecc. come un coltello. Lo stesso oggetto, caratteristiche differenti: alcune indipendenti dall'osservatore, alcune relative all'osservatore.

È tipico delle scienze naturali che esse si occupino delle caratteristiche indipendenti dall'osservatore – come la forza, la massa, il legame chimico ecc. – ed è tipico delle scienze sociali che esse si occupino delle caratteristiche relative all'osservatore, come i soldi, la proprietà, il matrimonio, il governo. Di solito, la psicologia sta a metà. Alcune parti della psicologia si occupano delle caratteristiche relative all'osservatore, ma la psicologia cognitiva, la parte che costituisce il nucleo della scienza cognitiva, si occupa delle caratteristiche indipendenti dall'osservatore come la percezione e la memoria.

Ovunque vi sia una caratteristica relativa all'osservatore, come l'essere un coltello o l'essere dei soldi, vi devono essere alcuni agenti che usano o trattano le entità in questione come coltelli o soldi. Ora, e questo è un punto importante, sebbene i soldi e i coltelli siano relativi all'osservatore, il fatto che gli osservatori trattino certi oggetti come denaro o coltelli non è relativo all'osservatore; esso è indipendente dall'osservatore. È intrinseco e mi riguarda il fatto che io tratti questo oggetto come un coltello, anche se il fatto che questo oggetto sia un coltello esiste solo in relazione a me e agli altri osservatori. Gli atteggiamenti degli osservatori riguardo a quali entità soddisfino le descrizioni relative all'osservatore non sono di per se stessi relativi all'osservatore.

Questo avviene perché le spiegazioni della scienza sociale possono soddisfare il vincolo di realtà causale anche se le caratteristiche a cui si richiamano sono caratteristiche relative all'osservatore. Così, per esempio, se dico “la crescita dei tassi di interesse americani ha causato una crescita del valore del dollaro nel cambio con la sterlina”, ciò è una spiegazione *causale* perfettamente legittima, anche se sterline, dollari e tassi di interessi sono tutti relativi all'osservatore. I meccanismi causali operano in tale spiegazione anche se lo fanno attraverso gli atteggiamenti degli investitori, dei banchieri, degli agenti di cambio, degli speculatori ecc. Sotto questo riguardo l'aumento del valore del dollaro non è come l'aumento della pressione del gas nel momento in cui questo viene scaldato. L'aumento della pressione del gas è indipendente dall'osservatore, l'aumento del valore del dollaro è dipendente dall'osservatore. Ma la spiegazione in entrambi i casi può essere una spiegazione causale. La differenza sta nel fatto che la spiegazione dei fenomeni relativi all'osservatore fa riferimento implicito agli agenti umani.

La terza distinzione è un'applicazione della seconda. Si tratta della distinzione tra intenzionalità *intrinseca* od originale e intenzionalità *derivata*. Se al momento sono in uno stato di sete o di fame, l'intenzionalità del mio stato è intrinseca a tali stati – entrambi implicano desideri. Se descrivo questi stati proferendo enunciati del tipo “ho sete” o “ho fame”, gli enunciati sono anche intenzionali perché soddisfano condizioni di verità. Ma l'intenzionalità degli enunciati non è intrinseca a essi in quanto sequenze sintattiche. Tali enuncia-

ti traggono il loro significato dall'intenzionalità di coloro che parlano italiano. Stati mentali come credenze, desideri, emozioni, percezioni ecc. hanno intenzionalità intrinseca; ma enunciati, mappe, quadri e libri hanno solo intenzionalità derivata. In entrambi i casi, l'intenzionalità è reale e ascritta in senso letterale, ma l'intenzionalità derivata deve essere derivata dall'intenzionalità originaria o intrinseca di effettivi agenti umani o animali.

Voglio che questa distinzione suoni ovvia, perché credo che lo sia. E credo anche che essa sia un caso speciale della distinzione ugualmente ovvia tra relatività rispetto all'osservatore e indipendenza dall'osservatore. L'intenzionalità derivata è relativa all'osservatore, l'intenzionalità intrinseca è indipendente dall'osservatore.

Vi sono, inoltre, ascrizioni intenzionali che non ascrivono nessuno di questi tipi di intenzionalità. Si tratta tipicamente delle ascrizioni metaforiche o come-se. Diciamo cose come "il mio prato ha sete perché c'è siccità", o "la mia auto ha sete perché consuma molta benzina". Sostengo che queste sono affermazioni metaforiche inoffensive di poco interesse filosofico. Esse significano, grossolanamente, che il mio prato o la mia auto si trovano in una situazione simile a, e si comportano in qualche modo come, un organismo che ha sete in senso letterale.

Una siffatta intenzionalità come-se non dovrebbe essere confusa con l'intenzionalità derivata. L'intenzionalità derivata è vera e propria intenzionalità a tutti gli effetti ma è derivata dall'intenzionalità intrinseca degli effettivi agenti intenzionali come coloro che parlano una determinata lingua. Per tale ragione, essa è relativa all'osservatore. Ma l'intenzionalità come-se non è affatto intenzionalità. Quando dico di un sistema che esso ha intenzionalità come-se, ciò non gli attribuisce intenzionalità. Dice semplicemente che il sistema si comporta come se avesse intenzionalità, anche se di fatto non ne ha.

Per riassumere queste distinzioni: bisogna distinguere tra comportamento guidato da regole e comportamento descritto da regole. Bisogna distinguere le caratteristiche indipendenti dall'osservatore da quelle relative all'osservatore. Inoltre, bisogna distinguere l'intenzionalità indipendente dall'osservatore (o intrinseca) dall'intenzionalità dipendente dall'osservatore (derivata) e dall'intenzionalità come-se.

5. INFORMAZIONE E INTERPRETAZIONE

Voglio ora applicare queste distinzioni al modello dell'elaborazione di informazioni della spiegazione cognitiva. Sosterrò che se il modello alla Marr deve avere forza esplicativa, il comportamento da spiegare per mezzo delle regole dell'elaborazione di informazioni deve essere guidato da regole e non solo descritto da regole. È possibile soddisfare tale condizione soltanto se l'informazione è intrinseca o indipendente dall'osservatore. Per fare la distinzione

tra Ludwig e il sasso che cade, dobbiamo dimostrare che Ludwig sta effettivamente seguendo una regola e ciò può avvenire solo perché egli ha un contenuto intenzionale intrinseco appropriato. La difficoltà del modello classico può adesso essere esposta in una forma preliminare. Ogni nozione chiave nel modello è relativa all'osservatore: *informazione*, *rappresentazione*, *sintassi*, *simbolo* e *computazione*, in quanto usate tipicamente nella scienza cognitiva, sono tutte nozioni relative all'osservatore, e questo ha come conseguenza che il modello classico nella sua forma attuale non può soddisfare il vincolo di realtà causale. Proverò nel prosieguo a esporre ciò in maniera più precisa.

Addentriamoci in queste nozioni, cominciando dai "simboli" e dalla "sintassi". Assumo come cosa ovvia che un segno o una forma o un suono siano simboli o enunciativi o altri congegni sintattici relativi solo a quegli agenti che assegnano a essi un'interpretazione sintattica. E invero, sebbene sia meno ovvio, penso sia anche vero che un'entità possa avere un'interpretazione sintattica solo se ha anche un'interpretazione semantica, perché i simboli e i segni sono elementi sintattici relativi solo a quei significati che essi hanno. I simboli devono simboleggiare qualcosa e gli enunciativi devono significare qualcosa. Simboli ed enunciativi sono in effetti entità sintattiche, ma l'interpretazione sintattica richiede una semantica.

Quando arriviamo alla "rappresentazione" la situazione è un po' più ingarbugliata. Una rappresentazione può essere o relativa all'osservatore o indipendente dall'osservatore. Mappe, diagrammi, quadri ed enunciativi sono tutti rappresentazioni e sono tutti relativi all'osservatore. Credenze e desideri sono rappresentazioni mentali e sono indipendenti dall'osservatore. Per di più un animale può avere rappresentazioni mentali quali credenze e desideri senza avere per nulla a che fare con entità sintattiche o simboliche. Quando Ludwig vuole mangiare o bere, per esempio, egli non ha affatto bisogno di usare alcun simbolo o enunciato per realizzare i suoi desideri canini. Egli ha soltanto fame o sete. La parte ingarbugliata deriva dal fatto che talvolta credenze e desideri indipendenti dall'osservatore fanno uso di enunciativi ecc., che sono relativi all'osservatore.

Invero alcuni filosofi hanno detto che tutte le credenze e tutti i desideri sono "atteggiamenti proposizionali" nel senso che sono atteggiamenti verso proposizioni o enunciativi o qualche altra forma di rappresentazione. Un tempo pensavo che questo fosse un errore innocuo, ma non lo è. Se credo che Clinton sia il presidente degli Stati Uniti, ho in effetti un atteggiamento verso Clinton, ma non verso un enunciato o una proposizione. L'enunciato "Clinton è il presidente degli Stati Uniti" è usato per esprimere la mia credenza e la proposizione che Clinton è il presidente degli Stati Uniti è il contenuto della mia credenza. Ma non ho alcun atteggiamento verso l'enunciato o la proposizione. Invero la proposizione, costruita come creduta, è proprio identica alla mia credenza. Non è l'oggetto della credenza.

La dottrina degli atteggiamenti proposizionali è un errore pericoloso per-

ché conduce la gente a postulare un complesso di entità nella testa, le rappresentazioni mentali, e avere una credenza o un desiderio si suppone sia avere un atteggiamento verso una di queste entità simboliche, simili a enunciati. Il punto per i nostri scopi è che le rappresentazioni mentali intrinseche quali le credenze e i desideri (stati intenzionali, come preferisco chiamarli) non richiedono espedienti rappresentativi, congegni sintattici, per esistere. E dove c'è un congegno sintattico, il congegno sintattico, essendo dipendente dall'osservatore, eredita il suo status sintattico e semantico dal contenuto intenzionale intrinseco della mente e non viceversa. Il punto cruciale per questa discussione è che tutte le entità sintattiche sono relative all'osservatore.

La distinzione tra caratteristiche del mondo indipendenti e dipendenti dall'osservatore si applica all'informazione. "Informazione" è chiaramente una nozione intenzionalistica, perché l'informazione è sempre informazione *su* qualcosa e tipicamente l'informazione è: le cose stanno così e così. L'"essere riguardo a qualcosa" (*aboutness*) in questo senso è la qualità che definisce l'intenzionalità, e il contenuto intenzionale di questo genere proposizionale è tipico dell'intenzionalità. Così non dovrebbe sorprendere che le distinzioni tra tipi differenti di ascrizioni intenzionali si applichino all'informazione. Perciò, se dico: "Conosco la strada per San Jose", ascrivo a me stesso l'informazione che non dipende da alcun osservatore. Essa è intrinseca o indipendente dall'osservatore. Se dico: "Questo libro contiene l'informazione su come raggiungere San Jose", il libro contiene in senso letterale l'informazione, ma l'interpretazione di ciò che è scritto nel libro in quanto informazione dipende dagli interpreti. L'informazione è dipendente dall'osservatore.

Vi sono anche ascrizioni di informazione come-se. Se dico: "Questi anelli del tronco contengono l'informazione sull'età dell'albero", questa non è un'ascrizione letterale di intenzionalità. Non vi è alcun contenuto proposizionale espresso dal legno. Ciò che in realtà dico, espresso alla lettera, è che una persona ben informata può inferire l'età dell'albero dal numero degli anelli, perché vi è un'esatta covarianza tra il numero degli anelli dell'albero e la sua età espressa in anni. Penso che con l'uso assai diffuso della nozione di "informazione", in particolar modo come risultato della teoria dell'informazione, molti direbbero che il tronco contiene in senso letterale l'informazione. Penso che essi credano di parlare in senso letterale quando dicono che il DNA contiene informazioni. Ciò è perfettamente ragionevole, ma è un significato differente di "informazione" quello che separa informazione e intenzionalità. Non vi è alcuna realtà psicologica per l'"informazione" nell'anello del tronco dell'albero o nel DNA. Essi non hanno né contenuto proposizionale né intenzionalità nel senso in cui i pensieri nella mia testa hanno un'intenzionalità originaria e le frasi nel libro hanno un'intenzionalità derivata.

Di questi tre tipi di ascrizione intenzionale soltanto l'informazione intrinseca è indipendente dall'osservatore.

A quale tipo di informazione si fa ricorso nelle teorie dell'elaborazione di informazioni proprie della scienza cognitiva? Ebbene, l'informazione come-se non ne fa. Se la spiegazione deve soddisfare il vincolo di realtà causale, si deve fare ricorso a qualche fatto informativo reale. Perché l'informazione derivata non dovrebbe soddisfare il vincolo di realtà? Dopo tutto, possiamo fornire veri e propri resoconti scientifici del flusso del denaro in economia; perché non resoconti scientifici del flusso delle informazioni nel sistema cognitivo, anche se l'informazione, come il denaro, è un fenomeno relativo all'osservatore? La risposta, in breve, è che, nel caso dell'economia, gli agenti che trattano determinati fenomeni fisici come il denaro sono parti dell'oggetto che stiamo studiando. Ma, in scienza cognitiva, se diciamo di fornire una spiegazione dei processi cognitivi dell'agente in base all'elaborazione di informazioni, non possiamo accettare una spiegazione nella quale l'elaborazione di informazioni dell'agente esiste solo in relazione alla sua intenzionalità, perché allora non avremmo spiegato l'intenzionalità dalla quale dipendono tutti i suoi processi cognitivi. In sintesi, saremo caduti nella fallacia dell'omuncolo. Se, dall'altro lato, pensassimo all'informazione come esistente in relazione soltanto a noi – gli osservatori – allora non avremmo soddisfatto il vincolo di realtà causale perché non avremmo identificato un fatto indipendente dalla teoria che spiega i dati che si suppone la teoria spieghi. Così, se le spiegazioni della scienza cognitiva andranno a soddisfare il vincolo di realtà causale, esse faranno ricorso all'informazione che è intrinseca all'agente, informazione che è indipendente dall'osservatore.

6. COMPUTAZIONE E INTERPRETAZIONE

Dunque, perché l'esigenza deve essere così forte? Perché non possiamo dire soltanto che il cervello si comporta come ogni altro computer? Diamo spiegazioni causali dei comuni computer, spiegazioni che soddisfano il vincolo di realtà causale ma che non ci obbligano a postulare un'intenzionalità intrinseca al computer.

La risposta è che la distinzione tra indipendente dall'osservatore e relativo all'osservatore si applica anche alla computazione. Quando addiziono 2 a 2 e ottengo 4 il calcolo aritmetico è intrinseco a me. È indipendente dall'osservatore. Quando, sul mio computer, digito "2 + 2" e ottengo "4", la computazione è relativa all'osservatore. I passaggi di stati elettrici sono tali – passaggi di stati elettrici – fino a quando un interprete non li interpreta come una computazione. La computazione non è intrinseca al silicio, né alla tensione elettrica. Io e gli altri come me siamo gli omuncoli del computer. Così, se diciamo che il cervello sta eseguendo la computazione, bisogna dire se essa è relativa o indipendente dall'osservatore. Se è indipendente dall'osservatore allora dobbiamo postulare un omuncolo dentro il cervello che sta effettivamente manipolando

simboli si da eseguire la computazione, proprio come io sto coscientemente manipolando i numeri arabi nel momento in cui addiziono 2 a 2 per ottenere 4. Se diciamo che essa è relativa all'osservatore, allora supponiamo che qualche osservatore esterno stia fornendo un'interpretazione computazionale degli impulsi neurali.

Ritengo che questo ultimo punto sia chiaro se ci si pensa su, ma non tutti lo trovano tale e perciò lo esaminerò un po' più avanti. Non riusciamo a vedere la relatività rispetto all'osservatore dell'ascrizione computazionale perché pensiamo che, dal momento che la computazione è tipicamente matematica e crediamo anche che il mondo soddisfi certe descrizioni matematiche in maniera indipendente dall'osservatore, in qualche modo ne deve seguire che la computazione è indipendente dall'osservatore. Comunque, c'è una sottile e nondimeno importante distinzione tra l'indipendenza dall'osservatore di certi fatti descritti matematicamente e la relatività rispetto all'osservatore della computazione che utilizza quei fatti. Si consideri l'esempio che ho fornito precedentemente del sasso che cade da un dirupo. Il sasso soddisfa la legge $S = 1/2 gt^2$, e quel fatto è indipendente dall'osservatore. Ma, si noti, possiamo trattare il sasso come un computer, se vogliamo. Supponiamo di voler calcolare l'altezza del dirupo. Conosciamo la regola e conosciamo la costante gravitazionale. Tutto ciò di cui abbiamo bisogno è un cronometro. Possiamo allora usare il sasso come un semplice analogo di un computer per calcolare l'altezza del dirupo.

Allora qual è la differenza tra il primo caso, in cui il sasso è proprio un sasso ed è descritto da una regola, e il secondo caso in cui il sasso è un computer che esegue una computazione, la quale implementa esattamente la stessa regola? La risposta è: al secondo caso abbiamo assegnato – cioè, c'è un'assegnazione relativa all'osservatore di – un'interpretazione computazionale. Ma ciò che è vero per il sasso è vero per ogni computer. Ciò che è particolare riguardo al sasso è che la legge di natura e l'algoritmo implementato sono la stessa cosa. In un computer commerciale utilizziamo le leggi di natura per assegnare altri algoritmi a processi elettronici: per l'addizione, la sottrazione, l'elaborazione di testi ecc. Ma il principio generale è questo: non possiamo ricorrere all'analogia tra il computer e il cervello per giustificare il carattere particolare del modello tripartito così come applicato al cervello, perché qualcosa è un computer solo in relazione a un'interpretazione computazionale.

Ciò che ho cercato di dimostrare con la parabola del sasso che cade è che la stessa descrizione matematica può essere considerata sia come una descrizione di un processo indipendente dall'osservatore, sia come una computazione relativa all'osservatore. Si tratta proprio di un fatto riguardante il sasso che cade in conformità alle leggi della fisica. Al riguardo non vi è nulla di relativo all'osservatore. Ma se consideriamo quel fatto computazionalmente, se

pensiamo al sasso come un qualcosa che esegue una computazione, allora una siffatta computazione esiste solo in relazione a noi.

Penso che si possa chiarire questo aspetto fornendo un esempio più semplice. Se è un fatto che vi siano tre mucche in un campo e due mucche in quello accanto, entrambi sono fatti indipendenti dall'osservatore. Se decido però di usare questi fatti allo scopo di effettuare un calcolo matematico e, nel contare le mucche, aggiungo 3 a 2 ottenendo 5, il processo computazionale di addizione non è qualcosa di intrinseco alle mucche nel campo. Il processo di addizione è un processo che io effettuo usando le mucche come la mia addizionatrice.

Ora, ciò che è vero per il sasso e per le mucche nel campo è vero per la computazione, generalmente. Se sto facendo coscientemente aritmetica, quella computazione è intrinseca. Se un'addizionatrice tascabile sta facendo aritmetica, la computazione è relativa all'osservatore. Vale la pena di osservare, incidentalmente, che con il passare degli anni la parola "computer" ha cambiato di significato.

Quando Turing scrisse il suo famoso articolo nel 1950, la parola "computer" significava "persona che calcola". Ecco perché Turing intitolò l'articolo *Macchine calcolatrici e intelligenza* e non *Computer e intelligenza*. "Computer" significava: persona che calcola. Oggigiorno, la parola "computer" ha cambiato il suo significato da fenomeno indipendente da un osservatore a fenomeno relativo a un osservatore. "Computer" adesso si riferisce a una classe di artefatti. Questo mutamento del significato di "computer", come il mutamento del significato di "informazione", ha contribuito a rendere confusa la distinzione tra intenzionalità intrinseca e altri tipi di fenomeni, e ha contribuito a favorire le confusioni che sto cercando qui di chiarire.

7. ELABORAZIONE DI INFORMAZIONI NEL CERVELLO

La questione cruciale per il modello classico può essere ora espressa con più precisione. Quale fatto corrisponde esattamente all'affermazione che vi è nel cervello un livello algoritmico di elaborazione di informazioni? E quale fatto corrisponde esattamente all'affermazione che tutto ciò che succede a questo livello si riduce a un livello di elaborazione primitiva che consiste interamente nella manipolazione di simboli binari? E questi processi computazionali di informazione sono indipendenti dall'osservatore o relativi all'osservatore?

Come primo passo domandiamoci come la pensano gli stessi che propongono questo modello. La risposta a questa domanda non è così chiara come dovrebbe essere ma penso che sia di questo tipo. A questo livello, il cervello lavora come un comune computer commerciale. Proprio come vi sono simboli nel computer, che portano informazioni, così vi sono enunciati nella testa

che portano informazioni. Proprio come il computer commerciale è un congegno per l'elaborazione di informazioni, così lo è il cervello.

Questa risposta è inaccettabile. Come abbiamo già visto, nel computer commerciale i simboli, gli enunciati, la rappresentazione, l'informazione e la computazione sono tutti relativi all'osservatore. Parlando dal punto di vista intrinseco, il computer commerciale è un circuito elettronico complicato. Affinché il computer commerciale soddisfi il vincolo di realtà causale dobbiamo ricorrere ai programmatori esterni, ai progettisti e agli utilizzatori che assegnano un'interpretazione all'input, ai processi intermedi, e all'output. Per il computer commerciale, noi siamo gli omuncoli che danno senso all'intera operazione.

Questo tipo di risposta non può funzionare per Ludwig perché qualunque altra cosa egli sia, è un agente intenzionale cosciente che prova a fare qualcosa, cerca di afferrare al volo una palla da tennis; e tutto ciò gli è intrinseco, niente di ciò è relativo all'osservatore. Vogliamo sapere come egli funziona realmente, intrinsecamente, non proprio che tipo di posizione potremmo adottare nei suoi confronti o quali interpretazioni computazionali potremmo imporgli.

Dunque, perché Ludwig non può computare in modo intrinseco? Perché non può eseguire algoritmi in modo inconscio nella maniera in cui io eseguo coscientemente un algoritmo per risolvere divisioni complesse? Possiamo dire che lo fa, ma se lo facciamo abbiamo abbandonato il modello, perché ora il meccanismo causale esplicativo non è l'algoritmo, ma l'agente mentale che all'interno sta intenzionalmente seguendo i passaggi dell'algoritmo. Questa risposta, in sintesi, rimanda alla fallacia dell'omuncolo. Noi non spieghiamo il comportamento di Ludwig che intenzionalmente cerca di afferrare al volo la palla per mezzo di un algoritmo se dobbiamo ricorrere al suo comportamento consistente nell'eseguire intenzionalmente la computazione della traiettoria parabolica, e spiegarlo successivamente con il suo comportamento consistente nel seguire intenzionalmente milioni di passaggi binari. Perché ora il meccanismo esplicativo del suo sistema rimane la sua irriducibile intenzionalità. L'idea del modello era che l'informazione nel sistema passasse attraverso le operazioni computazionali sulla sintassi. La semantica procede da sola. Ma in base a questa analisi è la sintassi che procede da sola. L'intenzionalità intrinseca dell'agente fa tutto il lavoro. Per vedere questo aspetto si noti che la spiegazione psicologica del mio eseguire divisioni complesse non è l'algoritmo, ma la mia *padronanza* dell'algoritmo e il mio seguire *intenzionalmente* i passaggi dell'algoritmo.

La conclusione per il modello classico può essere espressa nella forma di un dilemma: o le nozioni cruciali sono considerate in un senso relativo all'osservatore o lo sono in un senso indipendente dall'osservatore. Se lo sono nel senso relativo all'osservatore, allora la spiegazione fallisce perché non riesce a

soddisfare il vincolo di realtà causale. Se lo sono nel senso indipendente dall'osservatore, allora fallisce a causa della fallacia dell'omuncolo. È l'omuncolo che sta lavorando. Si tratta di scegliere tra un omuncolo esterno (relativo all'osservatore) e un omuncolo interno (indipendente dall'osservatore). Nessuna delle due opzioni è accettabile.

8. SEGUIRE REGOLE PROFONDE INCONSCIE

Penso che un modo per presentare il mio argomento sarebbe quello di offrire una prova effettiva che convinca del contrario. Vi sono esempi convincenti e certi del seguire regole computazionali profonde inconscie?

Ho sostenuto altrove che una determinata forma aspettuale richiede l'accessibilità alla coscienza almeno in linea di principio. In molti casi, per esempio in quello relativo alla "visione cieca", il contenuto in effetti non è accessibile alla coscienza, ma precisamente per tale ragione noi consideriamo questi casi patologici, dovuti a deficit, a repressione ecc. Non tornerò qui di nuovo su quell'argomento ma cercherò di porre una domanda differente: vi sono esempi certi del seguire regole profonde inconscie?

Se avessimo alcuni esempi convincenti, allora avremmo pochi dubbi a proposito del principio generale. Se potessimo essere d'accordo sul fatto che vi sono casi riguardanti il seguire regole in questo senso tecnico, che partono dalla nostra nozione ordinaria di senso comune riguardante il seguire regole, e se potessimo essere inoltre d'accordo sul fatto che queste spiegazioni hanno un vero e proprio potere esplicativo causale, allora avremmo almeno un buon inizio di una giustificazione per una strategia generale della scienza cognitiva volta a postulare tali spiegazioni mediante regole profonde inconscie. I due esempi che mi sono stati presentati sono il funzionamento del *modus ponens* e le altre regole logiche e, in secondo luogo, il funzionamento del riflesso vestibolo-oculomotore. (C'è un non so che di ironico circa il riflesso vestibolo-oculomotore – rvo – perché l'avevo presentato come qualcosa che ritenevo un esempio ovvio di un caso che sembrava soddisfare in superficie il vincolo di realtà causale, laddove era ovvio che non lo era.)⁴

Considererò ciascuno di essi in successione. La gente chiaramente è capace di fare inferenze logiche. Lo fa, secondo la descrizione, seguendo regole di cui è totalmente inconsapevole e che non potrebbe mai formulare senza assistenza professionale. Così, per esempio, la gente è capace di fare inferenze nella forma del *modus ponens*, e perciò segue la regola del *modus ponens*, anche se non potrebbe formulare la regola del *modus ponens* e, inverosimilmente, non ha il concetto di *modus ponens*.

4. Searle, *La riscoperta della mente*, cit., pp. 253-6.

Dunque, mettiamolo alla prova e vediamo come funziona. Ecco una tipica inferenza che utilizza il *modus ponens*. Prima delle elezioni del 1996 credevo che se Clinton avesse conquistato lo Stato della California avrebbe vinto le elezioni. Avendo esaminato i risultati elettorali in California, sono giunto alla conclusione che Clinton avrebbe conquistato la California, così ho inferito che egli avrebbe vinto le elezioni. Ora, come ho fatto questa inferenza?

Ebbene, la spiegazione della scienza cognitiva sarebbe: quando hai fatto l'inferenza, in effetti stavi seguendo una regola inconscia. Questa è la regola del *modus ponens*, la regola che dice che se poni delle premesse nella forma "p", e "se p allora q", allora puoi validamente inferire "q". Mi sembra comunque che, in casi come questi, la regola non giochi affatto un ruolo esplicativo. Se io credo che Clinton conquisterà la California e credo che se egli conquista la California vincerà le elezioni, ciò è già sufficiente per mettermi in grado di inferire che vincerà le elezioni. La regola non aggiunge nulla alla spiegazione della mia inferenza. La spiegazione dell'inferenza è che io posso vedere che la conclusione segue dalle premesse. Ma la conclusione segue dalle premesse soltanto perché essa esemplifica la regola del *modus ponens* – deriva la sua validità dal *modus ponens*? Credo che la risposta a queste domande sia: no. Il *modus ponens*, costruito come una regola sintattica computazionale, è semplicemente un modello che noi usiamo per descrivere inferenze che sono valide *indipendentemente*. Non seguiamo la regola del *modus ponens* allo scopo di fare l'inferenza. Piuttosto, rendiamo valida l'inferenza, e il logico può formulare la cosiddetta regola del *modus ponens* per descrivere un infinito numero di siffatte inferenze valide. Ma le inferenze non derivano la loro validità dal *modus ponens*. Piuttosto, il *modus ponens* deriva la sua validità dalla validità indipendente delle inferenze. Pensare altrimenti conduce al paradosso di Lewis Carroll⁵. Così, mi sembra, il *modus ponens* non gioca affatto un ruolo esplicativo in un'inferenza del tipo di quella che ho appena descritto.

Ma che dire delle inferenze puramente formali provate solo in via teorica? Supponiamo che io abbia un gruppo di simboli e inferisca da "p" e "p → q" "q". Ora mi sembra che una volta che abbiamo sottratto il contenuto semantico dalle proposizioni, c'è in realtà un ruolo per la regola del *modus ponens*. Ma allora proprio perché c'è una siffatta regola, noi non stiamo più parlando di inferenze valide come parte dei processi cognitivi umani. Stiamo parlando di un analogo formale di queste inferenze valide, entro un sistema formale provato solo in via teorica. Cioè, se ti viene data una regola che dice che tutte le volte che hai simboli della forma: "scarabocchio macchia schiribocchio", seguito da "scarabocchio", puoi scrivere "schiribocchio", quella è una vera e propria regola. Essa ti dice ciò che puoi fare in certe circostanze e presenta

5. L. Carroll, *What the Tortoise Said to Achilles*, in "Mind", 4, 1895, n. 14, pp. 278-80.

tutte quelle caratteristiche che ho descritto come tipiche del comportamento diretto da regole, o delle spiegazioni mediante regole, singolarmente considerate. Ma ciò non è precisamente il funzionamento della regola del *modus ponens* nel ragionamento ordinario. Per evidenziare in maniera precisa questo punto, se pensiamo al *modus ponens* come a una descrizione effettiva del funzionamento dei contenuti mentali, allora il *modus ponens* non gioca alcun ruolo esplicativo nelle inferenze valide. Se lo pensiamo come una regola provata solo in via teorica che descrive operazioni su simboli privi di senso, allora invero esso gioca un ruolo, ma il suo ruolo non è quello di spiegare come noi effettivamente facciamo inferenze nei processi cognitivi ordinari, bensì come possiamo rappresentare la struttura formale o sintattica di quelle inferenze in sistemi creati artificialmente.

Passo ora al riflesso vestibolo-oculomotore. Sembra che si stia inconsciamente seguendo la regola “muovi i bulbi oculari secondo movimenti eguali e opposti a quelli della testa”, quando in effetti non stiamo seguendo alcuna regola di tal fatta. Vi è un complesso meccanismo riflesso nel cervello che produce questo comportamento. Pensavo che la questione fosse ovvia, ma non è così. Recentemente alcuni dei miei critici hanno detto che vi sono anche stati computazionali sottostanti l’ambito della coscienza intrinseci al sistema che stanno a un livello più sottile rispetto alla regola che ho appena enunciato. Martin Davies dice:

Un’altra maniera per descrivere l’rvo rimanda a un sistema nel quale ha luogo una certa elaborazione di informazioni, non dai movimenti della testa ai movimenti degli occhi secondo certe velocità, ma dalle rappresentazioni delle velocità del movimento della testa alle rappresentazioni delle velocità del movimento degli occhi. È solo contro lo sfondo di questo secondo tipo di descrizione che si pone ogni questione relativa al dar credito al sistema con la conoscenza tacita delle regole che rapportano la velocità della testa alla velocità degli occhi⁶.

Questa assunzione del “contenuto semantico” negli stati in entrata e in uscita è una condizione necessaria ma non sufficiente della conoscenza tacita di regole. La condizione sufficiente richiede che «le varie trasformazioni in entrata e in uscita che sono conformi alla regola dovrebbero avere la stessa spiegazione causale»⁷.

L’rvo soddisfa facilmente tale condizione, sicché risulta che l’rvo è un

6. M. Davies, *Reply: Consciousness and the Variety of Aboutness*, in C. MacDonald, G. MacDonald (eds.), *Philosophy of Psychology: Debates on Psychological Explanation*, Blackwell, Oxford 1995, p. 386.

7. *Ibid.*

caso di conoscenza tacita inconscia di regole ed è un caso di comportamento diretto da regole. Per sostenere ciò, Davies fornisce vari resoconti di descrizioni computazionali dell'rvo tratte da David Robinson, Patricia Churchland e Terry Sejnowski⁸. Egli erroneamente pensa che io stia sostenendo che le ascrizioni computazionali siano poco importanti. Ma questo non è il mio argomento. Il mio argomento verte intorno alla *realtà psicologica* delle ascrizioni computazionali. Non vedo alcuna ragione per trattare la descrizione computazionale dell'rvo in maniera differente dalla descrizione computazionale dello stomaco e di altri organi. La mia domanda è: c'è un livello causale distinto dal livello della neurofisiologia nel quale l'agente sta effettivamente eseguendo in maniera inconscia certi compiti computazionali, di elaborazione di informazioni allo scopo di muovere i suoi bulbi oculari? Non vedo nulla nel resoconto di Davies per supporre che l'aver postulato un siffatto livello soddisfi il vincolo di realtà causale. Che cosa riguardo ai nuclei vestibolari fa sì che essi eseguano specificamente operazioni mentali al livello dell'intenzionalità intrinseca? Non vedo alcuna risposta a questa domanda. Non è un'obiezione all'utilità dei modelli computazionali dell'rvo indicare che essi sono modelli di neurofisiologia, e non esempi di processi psicologici effettivi: essi sono al livello dell'elaborazione di informazione neuronale relativa all'osservatore, non dell'intenzionalità intrinseca. Una cosa è avere una descrizione computazionale di un processo, un'altra eseguire effettivamente un processo mentale di computazione.

9. CONCLUSIONE

Sulla base del resoconto che sto proponendo, le descrizioni computazionali giocano nella scienza cognitiva esattamente lo stesso ruolo che giocano in ogni altro ramo della biologia o nelle altre scienze naturali. Eccetto i casi in cui un agente sta effettivamente eseguendo in maniera intenzionale una computazione, la descrizione computazionale non identifica un separato livello causale distinto dal livello della struttura fisica dell'organismo. Quando fornisci una spiegazione causale, ti chiedi sempre quale fatto causale corrisponde all'affermazione che stai facendo. Nel caso delle descrizioni computazionali dei processi cerebrali profondi inconsci, i processi sono descritti da regole e non diretti da regole.

E ciò che funziona per la computazione funziona *a fortiori* per l'"elaborazione di informazioni". Puoi fornire una descrizione del cervello in termini di elaborazione di informazioni così come la puoi fornire dello stomaco o della combustione interna di un motore. Ma se questa "elaborazione di informazio-

8. P. S. Churchland, T. J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge (MA) 1993.

ni” deve essere psicologicamente reale, essa deve essere un tipo di informazione intrinsecamente intenzionalistico. Le spiegazioni della scienza cognitiva, usando l’inconscio profondo, tipicamente non riescono a soddisfare questa condizione.

Vorrei concludere questa discussione con una diagnosi di ciò che penso sia un errore. È molto difficile per gli esseri umani accettare forme di spiegazione non animistiche, non intenzionalistiche. Nella nostra cultura siamo giunti pienamente ad accettare tali spiegazioni soltanto nel XVII secolo. Le nostre forme paradigmatiche di spiegazione sono intenzionalistiche: mangio questo cibo perché ho fame, bevo quest’acqua perché ho sete, guido a sinistra perché questa è la regola del codice stradale. L’idea che vi siano spiegazioni meccaniche che non rinvino ad alcuna intenzionalità è un’idea molto difficile da afferrare. Una forma di animismo ancora sopravvive in certi progetti di ricerca nella scienza cognitiva. Il livello intermedio di Marr del seguire regole al livello sottostante quello della coscienza nel cervello è una forma di animismo. Orbene, siccome questi processi supposti non sono coscienti, e nemmeno accessibili alla coscienza in linea di principio, postuliamo che ci si comporti seguendo regole profonde inconsce. Questo è l’errore dell’animismo primitivo. Ora, questo è aiutato da un secondo errore: siamo ingannati dall’apparente intenzionalità dei computer, dei termostati, dei carburatori e degli altri sistemi funzionali che abbiamo progettato. Ci sembra evidente che questi sistemi presentino un livello intenzionalistico di descrizione. Invero, i manuali canonici della scienza cognitiva pongono la descrizione di livello intermedio del termostato fornita da Marr come se la spiegazione di livello algoritmico soddisfacesse evidentemente il vincolo di realtà causale. Ma ritengo sia chiaro che non lo fa. Nel caso del termostato abbiamo allestito sistemi fisici in modo che si comportino come se stessero seguendo regole computazionali. Ma la computazione intenzionale, che segue regole, del termostato è completamente relativa all’osservatore. È solo perché abbiamo progettato e usato questi sistemi che possiamo fornire spiegazioni intenzionalistiche. Ora, ciò che vale per il termostato vale per gli altri sistemi funzionali, come gli orologi, i carburatori e, soprattutto, i computer. Così, stiamo facendo due errori. Il primo è un errore consistente nel preferire l’animismo alle spiegazioni naturalistiche, il secondo riguarda l’incapacità di distinguere tra relatività rispetto all’osservatore e indipendenza dall’osservatore. In particolare, non riusciamo a distinguere i casi in cui abbiamo una vera e propria intenzionalità intrinseca opposta all’intenzionalità relativa all’osservatore. L’intenzionalità nei termostati, nei carburatori, negli orologi e nei computer è interamente relativa all’osservatore.

La situazione in cui ci troviamo è esattamente la seguente: immaginiamo delle automobili come fenomeni naturali; non sappiamo come esse funzionino. Saremmo tentati di pensare che molto di ciò che fanno è un’elabora-

zione computazionale di informazioni. Per esempio potremmo cercare di spiegare il sistema del contachilometri dicendo che esso computa la velocità dell'auto in chilometri orari a partire dall'input di informazioni circa la rotazione della ruota intorno a un asse al minuto. Potremmo anche calcolare l'algoritmo per mezzo del quale si rappresentano i giri per chilometro orario. Ma una siffatta spiegazione non ha alcuna realtà causale. Il meccanismo causale effettivo è che un piccolo generatore elettrico è attaccato a una ruota in modo tale che l'aumento dei giri produca l'aumento dell'elettricità generata. Un amperometro con un ago si muove più in alto o più in basso man mano che l'elettricità aumenta o diminuisce. Per quanto riguarda la causazione, questo è tutto. Parlando dal punto di vista intrinseco, non c'è computazione né informazione. In aggiunta alla fisica pura, la computazione e l'informazione sono tutte in noi. Sono relative all'osservatore. Non c'è nulla di male nella computazione e nell'informazione relative all'osservatore. Dopo tutto, questo è ciò per cui noi in effetti progettiamo il sistema del contachilometri. La cosa sbagliata consiste nel confondere l'attribuzione relativa all'osservatore della computazione e dell'elaborazione di informazioni con la spiegazione causale. Ora, la cosa difficile da vedere è che molte delle descrizioni intenzionalistiche dei processi cerebrali sono anche relative all'osservatore, e conseguentemente non ci forniscono una spiegazione causale. Allora, qual è il modello corretto per la spiegazione della scienza cognitiva? E, invero, come descriviamo molta dell'apparente razionalità dei processi cognitivi se non postuliamo il comportamento diretto da regole al livello intermedio di Marr? Per rispondere a ciò, mi sembra che dobbiamo ricordarci del modo in cui Darwin ha risolto un problema simile mostrando come l'apparente finalismo nella struttura delle specie possa essere spiegato senza postulare alcuna intenzionalità. Secondo uno stile darwiniano di spiegazione, consideriamo due livelli esplicativi al posto di uno. Invece di dire "il pesce ha la forma che ha al fine di sopravvivere nell'acqua", diciamo (1) il pesce ha la forma che ha a causa della sua struttura genetica e (2) i pesci che hanno quella forma sopravviveranno meglio dei pesci che non ce l'hanno. Si noti che la sopravvivenza funziona sempre nella spiegazione ma non è più un fine. È soltanto qualcosa che succede.

Ora, analogamente, non dovremmo dire "il bulbo oculare si muove perché sta seguendo la regola del riflesso vestibolo-oculomotore". Dovremmo dire che il bulbo oculare si muove a causa della struttura del sistema visivo – è solo un processo meccanico. Non vi è alcun *seguire* regole. La regola, comunque, *describe* il comportamento del bulbo oculare, e il bulbo oculare soddisfa quella descrizione per ragioni fondamentalmente darwiniane. I bulbi oculari che si comportano in quel modo produrranno un'immagine retinica più stabile, e gli organismi che hanno un'immagine retinica stabile hanno una maggiore probabilità di sopravvivere rispetto agli organismi che non ce l'hanno.

Analogamente, Ludwig non segue la regola della traiettoria parabolica; piuttosto cerca di raffigurarsi dove stia andando la palla e salta per mettere la sua bocca al punto esatto. Dopo molta pratica egli riesce piuttosto bene in ciò. Ha un'abilità nella coordinazione zampa-occhio che consente di descriverlo per mezzo della regola della traiettoria parabolica, ma egli non sta seguendo quella regola. I cani che riescono a sviluppare tale abilità hanno una maggiore probabilità di sopravvivere rispetto ai cani che non ce l'hanno – o almeno essi hanno una maggiore probabilità di afferrare al volo palle da tennis.