da Alfredo Paternoster, *Introduzione alla filosofia della mente* (2002)

Buona parte delle discussioni che animano la filosofia della mente contemporanea ruotano attorno al problema di come conciliare una visione materialistica del mondo con la natura apparentemente non materiale dei fenomeni mentali. Nella filosofia odierna l'espressione «materialismo» viene comunemente usata per denotare la tesi secondo cui l'intera realtà è costituita da massa-energia e da quant'altro è contemplato dalla fisica di base: particelle elementari, forze, onde. Userò pertanto indifferentemente «fisicalismo» e «materialismo» per riferirmi a tale prospettiva. Il fisicalista, beninteso, non intende sostenere che l'unica vera scienza della natura è la fisica, ma che questa è la scienza fondamentale, la più basilare, in quanto ogni processo chimico, per esempio, è riducibile, almeno in linea di principio, a processi fisici. Nella cornice teorica fisicalistica una pietra e una creatura dotata di mente non differiscono per la natura degli elementi costitutivi ultimi, ma per certe proprietà di organizzazione che investono il modo in cui tali elementi si combinano tra loro dando luogo a strutture di complessità crescente. Pertanto, quando si parla di materialismo o di fisicalismo in filosofia della mente, ciò che si intende è che le menti non sono «cose», o «sostanze», che fanno parte dell'arredo metafisico del mondo alla stessa stregua degli atomi, dei tavoli o degli organi di un corpo umano: le menti non sono nient'altro che il prodotto dell'attività del cervello. Niente cervelli, niente menti (né anime).

Il fisicalismo si è progressivamente imposto in opposizione alla dottrina del dualismo[1] ontologico[2], la tesi cartesiana secondo cui la mente costituisce un tipo di realtà a sé. Descartes pensava infatti che nel mondo ci fossero due tipi fondamentali di enti o, in un linguaggio più classico, di sostanze: la *res cogitans,* la sostanza pensante, e la *res extensa,* i corpi materiali. Le menti sono altro dal corpo e da esso indipendenti; potrebbero esistere in assenza del corpo.

Quest'ultima tesi appare oggi inverosimile alla grande maggioranza degli studiosi principalmente perché mal si accorda con alcuni capisaldi dell'immagine scientifica del mondo, come la teoria dell'evoluzione o la possibilità di conferma empirica (gioverà ricordare a questo riguardo che, nonostante le molte critiche che le vengono rivolte, l'immagine scientifica del mondo si è guadagnata autorevolezza e consenso con una molteplicità di indiscutibili successi). Ciò non significa, tuttavia, che il materialismo sollevi meno problemi filosofici del dualismo. Se infatti ciò che chiamiamo «mente» non è che il complesso delle attività del sistema nervoso, resta da spiegare come ciò sia possibile: come fa un mero conglomerato di cellule, per quanto di grandissime dimensioni, a produrre proprietà così sofisticate e in apparenza del tutto estranee al mondo fisico come il pensiero, l'esperienza, il senso dell'io? E come è possibile coniugare l'apparente libertà delle proprie scelte in un quadro teorico sostanzialmente meccanicistico? Se non si danno risposte a queste domande, il materialismo, più che proporre una chiara visione in positivo di che cosa è la mente e di quale rapporto essa intrattenga col corpo, si limita ad essere la mera negazione del dualismo.

Il fisicalismo sembra, ad esempio, avere le risorse per venire a capo dell'annosa incapacità del dualismo di giustificare la coordinazione, l'armonia di funzionamento, tra mente e cervello. Le soluzioni invocate fin dal Seicento, dalla ghiandola pineale all'armonia prestabilita o all'intervento occasionale di Dio, sono tutte esplicativamente inadeguate. A prima vista nel fisicalismo questo problema semplicemente scompare: non c'è niente da sincronizzare perché c'è solo il corpo. Ma a questo punto il dualista può a buon diritto reclamare una spiegazione di come uno *stato mentale,* per esempio un'intenzione, possa causare un'azione fisica. In entrambi i casi sono in gioco dei fenomeni inspiegati.

Pertanto, se è vero che si possono portare alcuni solidi argomenti contro il dualismo, è altrettanto vero che anche le diverse teorie materialistiche della mente sono esposte a gravi difficoltà. Così, non deve stupire più di tanto che ci siano tutt'oggi alcuni autorevoli studiosi che si proclamano dualisti, e persino nel campo scientifico, come testimoniano i casi di John Eccles e Wilder Penfield.

Gli autori che difendono il punto di vista dualista insistono tipicamente sulla peculiarità dell'aspetto *fenomenico* o *qualitativo* del mentale. Se pensiamo agli stati mentali in primo luogo come a stati di esperienza, se cioè ci si concentra su che cosa si prova in questa o quella situazione (la tonalità di quel rosso, l'intensità di quel mal di denti, l'irresistibilità di quel desiderio, ecc.), l'idea di una radicale differenza ontologica rispetto agli stati e agli oggetti del mondo fisico scaturisce abbastanza naturalmente. Gli stati mentali non possono essere semplicemente stati fisici perché sono caratterizzati da alcune proprietà radicalmente estranee, almeno in apparenza, al mondo fisico. Questo argomento cartesiano è stato tuttavia fondatamente criticato sulla base della considerazione che uno stesso oggetto può avere due o più descrizioni diverse e nondimeno essere lo stesso oggetto. I sintagmi «Monte Bianco» e «il monte più alto d'Europa» esprimono due prospettive affatto diverse, ma unica è la cosa che designano. Analogamente, gli stati mentali caratterizzati nei termini delle loro proprietà fenomeniche potrebbero essere mere prospettive su stati cerebrali perfettamente descrivibili nel linguaggio delle neuroscienze; «mal di denti», per esempio, potrebbe essere un'espressione del linguaggio ordinario che in realtà denota la stimolazione di alcune fibre nervose, le fibre C. A dispetto di questa replica, tuttavia, l'impressione di un'irriducibile alterità dell'esperienza resiste e la questione del dualismo riaffiora almeno come problema di spiegare in che modo il corpo produce l'esperienza: per confutare completamente il dualismo si vorrebbe una teoria scientifica che descrivesse che cosa è l'esperienza *in prima persona*.

Più in generale, il problema che si pone al fisicalista è quello di chiarire quale ruolo - se ne ha uno - dobbiamo attribuire alla mente in una prospettiva metafisica che contempla solo l'esistenza del corpo. Di che cosa stiamo parlando, esattamente, quando usiamo la parola «mente»? A prima vista si direbbe che le opzioni disponibili al fisicalista siano di due tipi: o *ridurre* la mente a qualcosa di ontologicamente non sospetto[3], identificandola cioè con qualcosa di materiale, o *eliminare* la mente, negando in modo puro e semplice la sua esistenza. In realtà, vedremo che il problema può essere affrontato in modo assai più articolato, fornendo un tipo di soluzione che non può essere sottomesso a questo schema un po' drastico.

---

[1] Teoria che distingue due tipi di "cose" (o sostanze) irriducibili (tipicamente: la dottrina cartesiana)

[2] *Onto-logia* è la scienza di ciò che è (in greco *òn*), inaugurata da Parmenide e proseguita con Platone e con la *Metafisica* di Aristotele (da cui il nome *metafisica* per indicare la stessa scienza filosofica). Una differenza ontologica è una differenza relativa al tipo di realtà considerato, per esempio quella tra una cosa e la sua immagine

[3] La *realtà* di un sasso parrebbe, ad es., meno sospetta di quella di un fantasma o di un drago.

La teoria dell'identità psico-fisica

Il modo più ovvio di interpretare la tesi secondo cui la mente non è nient'altro che il prodotto dell'attività cerebrale consiste verosimilmente nel dire che, se per ogni stato mentale esiste un preciso correlato neurologico, allora lo stato mentale è identico a tale correlato. Per esempio, avere mal di denti è avere una certa stimolazione delle fibre C, esattamente nello stesso senso in cui il calore è il moto molecolare o l'acqua è H_20. Analogamente, ogni volta che crediamo o desideriamo qualcosa, per esempio che ci sia della birra in frigorifero, ci troviamo in una determinata configurazione neuronale. Questa tesi, che prende il nome di *teoria dell'identità di tipo* (o *tipo-tipo*), è stata avanzata negli anni Cinquanta da Feigl, Place, Smart e ripresa, in una cornice teorica diversa, da Armstrong e in generale dalla fiorente scuola filosofica australiana.
Per comprendere il nome che è stato dato alla teoria, dobbiamo fare riferimento alla cruciale distinzione, forse la madre di tutte le distinzioni filosofiche, tra particolari ed universali o, in una terminologia desunta da Peirce, tra *occorrenze* (*tokens*) e *tipi* (*types*)[4]. Nella parola «gatto», per esempio, vi sono cinque occorrenze di lettere (particolari), ma solo quattro tipi di lettere (universali), perché vi sono due occorrenze della «t». Analogamente, quando si parla di stati mentali, possiamo riferirci a occorrenze o a tipi: il mal di denti che sto provando in questo momento, o il mal di denti in generale, che io e voi possiamo provare in diversi momenti. Generalmente il contesto non lascia dubbi su quale sia l'interpretazione da dare all'espressione «stato mentale», come occorrenza piuttosto che come tipo. Per esempio, quando diciamo che due stati mentali sono identici (o che uno stato mentale è identico a un altro) chiaramente ciò che si intende è che si tratta di due occorrenze dello stesso tipo. Si parla anche di proprietà mentali, per denotare tipi di stati mentali, e di eventi mentali, per denotare occorrenze di stati.
Dunque, come suggerisce il nome che è stato dato alla teoria, l'identità in questione non è soltanto tra uno stato mentale particolare (per esempio, il pensiero che ho in questo momento che è ora di andare a pranzo) e uno stato cerebrale particolare, bensì tra tutti gli stati mentali di un certo tipo e *tutti* gli stati cerebrali di un altro tipo, ovvero tra una certa proprietà mentale e una certa proprietà cerebrale. Ogni volta che io o voi pensiamo che è ora di andare a pranzo ci troviamo in uno stato cerebrale di uno stesso e unico tipo, ed è in questo senso che il tipo di stato mentale denotato da «credere che sia ora di andare a pranzo» è identico a un certo tipo di stato cerebrale.
La teoria dell'identità è un tipico esempio di dottrina riduzionistica. L'idea è che una caratterizzazione [descrizione] appropriata degli stati mentali debba passare attraverso una *riduzione interteorica*. La temperatura di un gas, per esempio, è identica all'energia cinetica media delle molecole che lo compongono. Questo genere di identità prende il nome di *legge-ponte*; un insieme di leggi-ponte consente di ridurre la teoria di alto livello a quella di basso livello, in questo caso di ridurre la termodinamica, una branca della macrofisica, alla meccanica statistica, che è microfisica. Analogamente si presume che ci siano leggi-ponte tra neurologia e psicologia così da poter ridescrivere uno stato postulato da una teoria psicologica di alto livello nei termini di proprietà di livello neuronale e, indirettamente, di livello fisico.
Queste considerazioni evidenziano come la nozione di riduzione interteorica comporti sempre un intreccio di due piani diversi, il piano ontologico (o metafisico) ed il piano epistemologico[5]. La riduzione in senso epistemologico è una *spiegazione* dei fenomeni descritti da una teoria T nei termini di una teoria di livello inferiore, più fondamentale, T'; si ha invece riduzione in senso ontologico quando l'insieme di fenomeni ridotto non presuppone l'esistenza di nessun altro *ente*[6] oltre a quelli coinvolti nei fenomeni riducenti. I sostenitori della teoria dell'identità contraggono, nel fare un'affermazione di portata chiaramente ontologica, anche un impegnativo debito epistemologico: ipotizzano l'esistenza di un insieme di leggi-ponte, ad oggi del tutto sconosciute, che connetterebbero psicologia e neuroscienza. In questo senso la tesi metafisica dell'identità psico-fisica può essere considerata alla stregua di un'ipotesi scientifica che guida un programma di ricerca.
Il maggior pregio della teoria dell'identità è evidente: rimuove qualsiasi appello a magici parallelismi o altri bizzarri meccanismi di interazione (tipo ghiandola pineale) caratteristici di questa o quella versione del dualismo. Gli stati mentali possono causare stati fisici perché sono stati fisici. Inoltre, in virtù della postulazione di leggi-ponte psico-fisiche, la teoria dell'identità promette di ricomprendere la psicologia all'interno delle scienze della natura.

L'obiezione della realizzabilità multipla. La tesi secondo cui *tutte* le occorrenze di un dato tipo mentale sono identiche a occorrenze di uno stesso tipo cerebrale è stata oggetto di una critica, dovuta a Putnam (1967), che ha minato la credibilità della teoria dell'identità. Il fatto che una proprietà mentale sia identica a una proprietà cerebrale comporta, per esempio, che tutte le volte che io e Simone crediamo che ci sia della birra in frigorifero ci troviamo nello stesso stato cerebrale, un'assunzione chiaramente molto vincolante. Se poi si considera che talora attribuiamo stati mentali anche ad animali, ne scaturisce l'assurda conclusione che due animali aventi cervelli diversi devono trovarsi in configurazioni cerebrali identiche ogni qual volta condividono uno stato mentale. Quando ho paura di essere divorato da uno squalo dovrei trovarmi nello stesso stato cerebrale di una sardina!

Altre obiezioni. Un'altra critica rivolta fin dall'inizio ai riduzionisti del mentale è che l'identità deve soddisfare il principio leibniziano dell'indiscernibilità degli identici:

$$\forall x \forall y((x=y) \rightarrow \forall F(Fx \leftrightarrow Fy)$$

per ogni x e y, se x è uguale ad y, allora, per ogni proprietà F, x ha F se e solo se y ha F.

[4] In termini platonici si tratta della distinzione tra fenomeni e idee.
[5] Se l'*onto-logia* è lo studio di ciò che è, l'*epistemo-logia* è lo studio della scienza stessa (*epistéme*), ossia la riflessione sul nostro modo di conoscere la realtà, distinto da questa realtà stessa.
[6] *Ente*, da *essere*, come *vedente*, da *vedere = ciò che è*, sinonimo di "cosa" e oggetto dell'ontologia.

Cioè: se due cose sono identiche, allora devono condividere tutte le proprietà. Ma è chiaro che ci sono proprietà caratteristiche di uno stato mentale - per esempio, le sue qualità fenomeniche - che non possono essere predicate di uno stato cerebrale (e, viceversa, ci sono proprietà chimiche o elettriche di uno stato cerebrale che non si predicano di uno stato mentale). Si noti come questa critica riprenda sostanzialmente l'argomento cartesiano dell'irriducibile alterità dell'esperienza. Come già accennato, tuttavia, a questa obiezione si può efficacemente replicare che è richiesta soltanto identità di riferimento, cioè di oggetto denotato, non identità sotto ogni aspetto: come «Monte Bianco» e «il monte più alto d'Europa» possono denotare lo stesso oggetto da due prospettive diverse, analogamente è possibile caratterizzare uno stesso tipo di stato in due modi diversi, con una descrizione mentale piuttosto che con una descrizione cerebrale.
Punta nella stessa direzione, ma in modo più sottile, l'assai discussa *critica di Kripke.* Supponiamo di aver trovato una correlazione attendibile tra il dolore e la stimolazione delle fibre C, per cui, in base alla teoria dell'identità, dolore = stimolazione delle fibre C. Ora, sostiene Kripke, le identità come «acqua = H_2O», sul cui modello è ricalcata l'identità psico-fisica, sono, se vere, vere di *necessità* (Premessa I); invece, l'identità tra il dolore e la stimolazione delle fibre C è chiaramente *contingente* (Premessa 2), poiché è concepibile che le fibre C siano stimolate senza che il soggetto provi una sensazione di dolore o che una sensazione di dolore insorga in modo diverso, per esempio tramite la stimolazione di un altro gruppo di fibre. Ne consegue che l'identità in questione non può essere vera perché non è necessaria (Conclusione).
L'argomento di Kripke è particolarmente spinoso perché entrambi i modi di respingerlo, attaccando la prima piuttosto che la seconda premessa, comportano difficoltà non lievi. Possiamo respingere la Premessa 1 negando che gli asserti di identità come «acqua = H_2O» siano necessari: l'acqua avrebbe potuto non essere H_2O, il dolore potrebbe non essere la stimolazione delle fibre C. Questa strategia è tuttavia non agevole da perseguire perché gli argomenti addotti da Kripke per giustificare lo statuto di necessità di certi asserti di identità sono abbastanza solidi.

[La tesi secondo cui gli asserti di identità come «acqua - H_20» sono necessari deriva da una teoria del riferimento dei nomi propri e di quelli di genere naturale, di sostanza che Kripke ha avanzato nel 1970, nota come *teoria della designazione rigida*. Per dare un'idea della teoria bisogna preliminarmente introdurre la nozione di mondo possibile. Un mondo possibile è un mondo (immaginario) in cui certi fatti sono andati diversamente che nel mondo reale, per esempio un mondo in cui il Presidente degli Stati Uniti è Al Gore e non George Bush. I mondi possibili sono cioè «costruiti» in base a una o più ipotesi controfattuali, relative a cose che avrebbero potuto andare diversamente da come sono andate di fatto. Ora, dire che certi nomi sono designatori rigidi equivale a dire che il loro riferimento è lo stesso in ogni mondo possibile; in altri termini non è possibile che un designatore rigido non denoti ciò che esso denota nel nostro mondo. Ne consegue che un'identità in cui compaiono due designatori rigidi, quali sono secondo Kripke «acqua» e «H_2O» è vera in ogni mondo possibile, cioè è vera di necessità, a differenza di un'identità come «Aristotele = Il maestro di Alessandro Magno», che è contingente perché l'espressione «Il maestro di Alessandro Magno» *non* è un designatore rigido. L'argomento alla base della teoria della designazione rigida è, almeno nel caso dei nomi propri, molto solido, e deriva dall'osservazione del funzionamento del ragionamento controfattuale ordinario, nel quale ascriviamo un'ipotetica proprietà a qualcosa o a qualcuno mantenendo fisso il riferimento del nome. Per esempio, quando diciamo «Se Aristotele fosse stato il maestro di Giulio Cesare...» o «Se l'acqua avesse un inebriante sapore di assenzio ... », stiamo facendo delle ipotesi controfattuali su quell'individuo particolare, Aristotele, che visse su questa terra, o sulla *nostra* acqua, l'acqua che c'è qui nel nostro mondo. Ne consegue che *in ogni altro mondo possibile,* come quello individuato dall'ipotesi che l'acqua abbia un inebriante gusto di assenzio, la parola «acqua» continua a denotare ciò che denota nel nostro mondo, *anche se le attribuiamo proprietà descrittive diverse.* Il lettore obietterà: ma così come si può immaginare che l'acqua abbia gusto di assenzio, si potrà ben immaginare che l'acqua non sia H_20; come può quindi l'identità in questione essere necessaria? Attenzione: in base alla teoria della designazione rigida, immaginare un mondo possibile in cui l'acqua non è H_20 è in effetti immaginare un mondo in cui ciò che noi chiamiamo «acqua», vale a dire, ciò che ha le proprietà superficiali dell'acqua, non ha una certa struttura chimica. Ma questo è come dire che stiamo immaginando un mondo in cui ciò che *sembra* acqua non è in realtà acqua; stiamo immaginando un mondo in cui ciò che esce dai rubinetti, si trova nei fiumi e nei laghi ecc. è una sostanza straordinariamente simile all'acqua, ma non è acqua (non è H_20) . La possibilità che l'acqua non sia H_20 è solo apparente, è un modo fuorviante di descrivere ciò che in effetti stiamo immaginando. Sono state mosse varie altre obiezioni alla teoria della designazione rigida, specialmente nel caso dei nomi di sostanza e specie naturale. Si è sostenuto, per esempio, che la teoria presuppone l'inaccettabile idea che le cose abbiano proprietà *essenziali.* Nondimeno, sarebbe difficile negare che la confutazione della teoria della designazione rigida, e con essa della prima premessa dell'argomento di Kripke, è un compito assai arduo, che richiede l'elaborazione di una teoria semantica alternativa che sia altrettanto efficace nel dar conto del funzionamento del discorso controfattuale]

In alternativa, possiamo prendere la strada di criticare la Premessa 2 mettendo in discussione l'affidabilità delle nostre intuizioni su casi controfattuali di questo genere. In realtà si potrebbe sostenere che non abbiamo alcuna intuizione chiara sulla possibilità di dissociare stimolazione delle fibre C e sensazione di dolore. L'identità *sembra* contingente ma è in effetti necessaria. Qui il problema è che l'onere della prova ricade su chi difende l'ipotesi che l'identità sia necessaria. In base a quali considerazioni dovremmo escludere la possibilità che, per esempio, una sensazione di dolore insorga senza una concomitante stimolazione delle fibre C? Per analogia col caso dell'acqua , dovremmo far vedere che questa è una possibilità solo *apparente,* che in realtà quello che stiamo immaginando è che qualcosa che ci *sembra* dolore, qualcosa che *percepiamo come* dolore, *non è* in effetti dolore (in quanto non è realizzato cerebralmente tramite la stimolazione delle fibre C). Ma questo è implausibile, *perché qualcosa che percepiamo come dolore, qualcosa che ci sembra dolore, è dolore*. *Il dolore è precisamente quello che percepiamo come tale*, la sensazione dolorosa; non c'è differenza fra provare (una sensazione di) dolore e avere un dolore.

Quale che sia il valore che si vuole attribuire agli argomenti contro la teoria dell'identità, si deve comunque riconoscere che allo stato attuale la scienza cognitiva - e la neuroscienza in particolare - è lontana dal poter offrire esempi di identità tra stati mentali e stati cerebrali, sebbene recenti ricerche di neuroscienza computazionale inducano a guardare al futuro con maggiore ottimismo. La teoria affida quindi a un futuro incerto le sue prospettive, anche se si deve ammettere che i successi delle riduzioni interteoriche in altri ambiti scientifici costituiscono un precedente favorevole. A volte le identità in questione sono ipotesi che guidano specifici programmi di ricerca; altre volte esse sono il punto d'arrivo di complesse risistemazioni teoriche innescate da scoperte empiriche di grande rilievo; i riduzionisti del mentale esemplificano il primo tipo di atteggiamento.

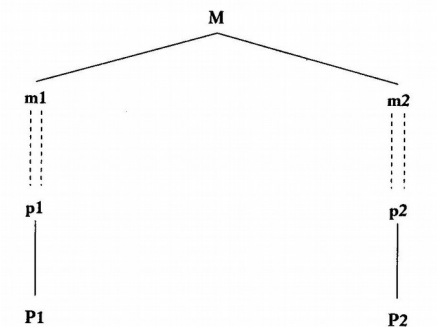
Il funzionalismo e il problema della causalità mentale

Il funzionalismo è la posizione filosofica riguardo alla natura della mente forse maggiormente in voga, o almeno così è stato per molti anni. Le ragioni della sua popolarità sono principalmente due. Da un lato, esso fornisce la cornice teorica più appropriata alle scienze della mente: la maggior parte degli scienziati cognitivi fa proprie, più o meno esplicitamente, le tesi fondamentali del funzionalismo. Dall'altro lato, esso consente di adottare una prospettiva *non* riduzionistica pur restando all'interno di un punto di vista fisicalistico.
L'idea fondamentale del funzionalismo è che uno stato mentale è il tipo di ente che è in virtù del suo svolgere una certa funzione, ovvero ricoprire un certo ruolo, all'interno dell'intera attività mentale di un agente. In una formula, gli stati mentali sono stati funzionali.
In termini generali, una proprietà di un sistema è funzionale se è individuata dal ruolo o, appunto, dalla funzione che certi elementi assolvono in quel sistema, prescindendo da come tale ruolo è realizzato fisicamente. Si tratta quindi di proprietà che investono l'organizzazione o la struttura logica di un sistema e non la sua composizione fisica. Un chiaro esempio di proprietà funzionale è quella di essere un centravanti, che un individuo soddisfa se e solo se svolge un certo ruolo all'interno di una squadra di calcio. Ma, per fare un caso più interessante, anche essere un cuore è una proprietà funzionale, nella misura in cui ciò che fa di un cuore l'organo che è non è la sua composizione fisica bensì un insieme di funzioni (pompare il sangue, assicurare nutrimento e ossigenazione dei tessuti, ecc.); e infatti esistono cuori artificiali. Lo stesso si può dire delle proprietà espresse dai termini che denotano artefatti (essere un martello, essere un carburatore, ecc.).
Applicata agli stati mentali, quest'idea appare intuitivamente molto plausibile. L'attività mentale può infatti essere vista come un insieme di complesse catene causali di stati. Queste relazioni causali possono essere ricondotte a tre tipi principali: quelle tra ambiente e stati interni (relazioni di *input* o *percettive,* per esempio una botta a un ginocchio mi fa avvertire una fitta dolorosa, oppure vedo un dolce in una vetrina e decido di entrare ad acquistarlo), quelle tra stati interni e stati interni (relazioni *cognitive in senso stretto,* come quando passo da una premessa a una conclusione in un ragionamento), quelle tra stati interni e comportamento (relazioni di *output* o *comportamentali*, per esempio un formicolio a una gamba mi fa cambiare posizione, oppure l'intenzione di prendere la parola mi fa alzare la mano). Pertanto i ruoli funzionali in questione sono ruoli causali; gli stati mentali sono cioè individuati nei termini del ruolo *causale* che essi svolgono. Il dolore, per esempio, può essere definito come ciò che è tipicamente causato da un danno arrecato al corpo o agli organi interni, causa uno stato di sofferenza più o meno intensa e la credenza che c'è una lesione in qualche parte del corpo, induce a comportamenti volti ad individuare l'origine del dolore e a cercare di alleviarlo in qualche modo. Analogamente, credere che stia piovendo è causalmente determinato da un corrispondente stato percettivo - vedo che fuori piove, o vedo la pioggia fuori - ed è a sua volta causa, congiuntamente al desiderio di non bagnarsi, dell'atto di prendere l'ombrello.
L'affermazione che gli stati mentali sono individuati dai loro ruoli causali suggerisce che gli stati mentali abbiano efficacia causale. In realtà non è necessario sposare questo punto di vista. Anzi, secondo il modo più comune di vedere la questione, gli stati mentali sono cause solo in quanto sono realizzati fisicamente. In altre parole, possiamo certamente dire che uno stato mentale occupa un ruolo causale, ma non che possiede, alla lettera, proprietà causali, perché un oggetto astratto non può avere proprietà causali. Sono gli stati fisici sottostanti ad essere causalmente efficaci, consentendo in tal modo allo stato mentale in questione di soddisfare il suo ruolo causale. In questo senso il funzionalismo è una forma di materialismo.
Tuttavia quello della *causalità mentale* resta uno dei problemi più spinosi correlati al funzionalismo. Proprio la seconda circostanza che ha favorito l'avvento del funzionalismo, cioè il suo dischiudere la possibilità di sposare una forma di materialismo *non riduzionistico*, ci permette di gettare luce su questo problema.
Non vi è dubbio che il funzionalismo sia emerso anche come reazione alla teoria dell'identità di tipo, insoddisfacente per le ragioni discusse sopra. Si tratta, ora, di misurare la capacità del funzionalismo di sostenere questo ruolo: presenteremo una metafisica della mente che si accompagna comunemente al funzionalismo e ne valuteremo la sua sostenibilità.
Il problema che stiamo per affrontare, quello della *causalilà mentale,* è attualmente uno dei più dibattuti; è infatti nei suoi termini che la filosofia della mente più recente ha reimpostato il classico problema mente-corpo.

Realizzabilità multipla e sopravvenienza

Benché il funzionalismo di per sé non comporti una determinata posizione riguardo alla relazione che sussiste tra stati mentali e stati cerebrali, di fatto si è accompagnato a una tesi, nota come «realizzabilità multipla», secondo cui ogni tipo di stato (= ogni proprietà) mentale può essere realizzato da una molteplicità di stati cerebrali di tipo differente. In questo modo possiamo dire che lo stato fisico in cui mi trovo quando provo dolore è diverso da quello (per esempio) di un cane o di una rana che provano dolore; e diventa possibile dar conto della verosimile ipotesi secondo cui, anche in una stessa persona, due stati di dolore provati in due istanti diversi possono corrispondere a una diversa configurazione neuronale. La realizzabilità multipla appare poi ancora più persuasiva per stati

mentali come credenze e desideri, visto che avere una certa credenza (o desiderio) sembra essere intuitivamente compatibile con una gamma molto ampia di diversi stati fisici.



Pertanto (cfr. figura) dal fatto che due stati mentali m_1 e m_2 siano occorrenze di uno stesso tipo M non possiamo trarre alcuna conclusione relativamente ai tipi fisici sottostanti; tutto quello che possiamo dire è che c'è uno stato cerebrale p_1, co-occorrente con m_1, e uno stato cerebrale p_2, co-occorrente con m_2, tali che p_1 e p_2 possono essere di tipo differente ($P_1 \neq P_2$). P_1 e P_2, o più propriamente, p_1 e p_2 sono diversi possibili realizzatori di M.

Si noti come questa formulazione lasci ancora aperta la possibilità di una forma di identità mente-cervello: si può ancora sostenere che ogni singola occorrenza di stato mentale è identica a un'occorrenza di stato cerebrale, ovvero che ogni evento mentale è identico a un evento cerebrale ($m_i = p_i$, per i = 1, 2, ...). Questa tesi è nota come *teoria dell'identità di occorrenza*, o *occorrenza-occorrenza* (*token-token identity*), ed è evidentemente contrapposta alla teoria dell'identità standard, che prevede un'identità tra *tipi* di stati mentali. La teoria dell'identità di occorrenza è il modo in cui molti filosofi funzionalisti concepiscono la relazione tra gli stati mentali e gli stati cerebrali. Infatti, poiché l'identità di tipo implica l'identità di occorrenza ma non viceversa, la teoria dell'identità di occorrenza corrisponde a una posizione più debole, meno vincolante.

Alla realizzabilità multipla si accompagna tipicamente, nella direzione opposta, anche la tesi seguente: fissato uno stato fisico di un certo tipo, risulta univocamente determinato lo stato mentale corrispondente; ne consegue che, se un organismo si trova in due istanti differenti t_1 e t_2 in due stati cerebrali identici, anche i corrispettivi stati mentali saranno identici. In una formula:

$$(S) \qquad (P_1 = P_2) \rightarrow (M_1 = M_2) \text{ [o, equivalentemente: } (M_1 \neq M_2) \rightarrow (P_1 \neq P_2)]$$

Identità di tipo fisico implica identità di tipo mentale (ma l'implicazione non vale nella direzione opposta) o, ciò che è lo stesso, una differenza (in tipo) di stato mentale implica una differenza fisica.

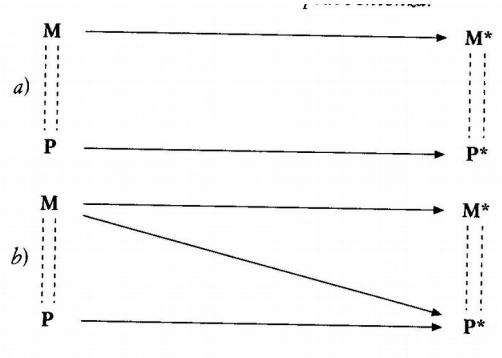
Questo tipo di relazione tra mentale e fisico è nota come *sopravvenienza psico-fisica.*

Il concetto di sopravvenienza, introdotto originariamente in filosofia morale, ammette in realtà diverse varianti. Ora ci limitiamo ad osservare che la sopravvenienza è comunemente considerata la nozione in cui trovano una sintesi tre diverse intuizioni relative al rapporto tra mentale e fisico: *covarianza*, *dipendenza* e *irriducibilità*.

Per *covarianza* si intende che qualunque differenza mentale deve avere riscontro in qualche differenza fisica; non possono esserci due stati identici sotto ogni aspetto fisico ma diversi per qualche aspetto mentale. t il nucleo logico fondamentale della nozione di sopravvenienza, quello catturato dalla (S). Per *dipendenza* si intende che gli stati sopravvenienti dipendono ontologicamente dagli stati fisici sottostanti, sia nel senso che, se non esistessero gli stati fisici, non esisterebbero nemmeno gli stati mentali, sia in quanto le caratteristiche dei secondi sono determinate da quelle dei primi. Con le parole di Kim, «il carattere mentale di una cosa è completamente determinato dalla sua natura fisica». Per *irriducibilità* si intende che, nonostante covarianza e dipendenza, il mentale conserva una sua autonomia esplicativa in quanto un *tipo* di stato o proprietà mentale non è in alcun modo riducibile: non c'è una proprietà fisica con cui identificare il mal di denti, o la credenza che Torino sia a nord di Roma. Ciò equivale a dire che, nella misura in cui le leggi correlano proprietà e non eventi, non ci sono leggi psico-fisiche, ad eccezione della (S) stessa.

Nel loro insieme, il funzionalismo e la sopravvenienza danno luogo a una prospettiva sulla mente che da un lato è coerente col fisicalismo e dall'altro salvaguarda il mentale come livello di spiegazione autonomo. E tuttavia questa autonomia ha qualcosa di sconcertante, se non di paradossale: non è forse tale una proposta teorica che da un lato rivendica la dipendenza ontologica del mentale dal fisico e dall'altro nega la possibilità stessa di leggi psico-fisiche? L'obiettivo di difendere il fisicalismo respingendo a un tempo il riduzionismo rischia di essere una strada molto stretta, così stretta che qualcuno dubita che sia coerentemente perseguibile. quanto cercheremo di appurare.

L'argomento dell'esclusione causale



Abbiamo visto che la metafisica della mente favorita dal materialismo non riduzionistico è basata sulle nozioni di realizzabilità multipla e di sopravvenienza. Questo tipo di soluzione al problema mente-corpo si esporrebbe tuttavia, secondo alcuni autori (per es. Kim) alla seguente grave difficoltà.

Siano *m* ed *m** due occorrenze rispettive di due diverse proprietà mentali M ed M*; per esempio, M è avere mal di testa e M* è avere l'intenzione di prendere un'aspirina. Siano *p* e *p** (rispettivamente di tipo P e P*) i realizzatori fisici di *m* e *m**. In base alla definizione di sopravvenienza, una qualsiasi occorrenza di P, per esempio *p*, è condizione sufficiente per un'occorrenza di M (lo stesso vale per P* ed M*). Siamo allora posti di fronte al seguente dilemma:

- Da un lato, secondo un'interpretazione intuitivamente plausibile del funzionalismo, saremo propensi a dire che M causa M* solo *in quanto* P causa P* (cfr. Fig. *a)*): è lo stato cerebrale che realizza il mal di testa a causare uno stato cerebrale che è alla base della disposizione a prendere l'aspirina. Si noti come la relazione causale sia definita fra *proprietà,* perché (almeno secondo Kim) l'esistenza di un nesso causale tra *m* ed *m** (così come tra *p* e *p**) deve essere espressione di una regolarità; il legame causale deve cioè sussistere per qualsiasi occorrenza di M e di M* (di P e P*). Ma se si interpreta il funzionalismo in questo modo, riconoscendo cioè un'autentica efficacia causale soltanto al realizzatori fisici, ne consegue l'*epifenomenismo* del mentale, cioè la tesi dell'irrilevanza causale degli stati mentali. Un epifenomeno si accompagna al fenomeno «vero» soggiacente come un'ombra a un corpo: c'è ma è ininfluente. Gli stati mentali sono mere ombre. L'epifenomenismo mette in discussione l'esistenza stessa degli stati mentali, almeno se si ritiene che qualunque cosa esistente debba avere qualche potere causale.

- Dall'altro lato, potremmo voler insistere sull'irriducibilità degli stati mentali, affermando che M causa M* in un senso genuino, non meramente «per procura». Ora, poiché M* dipende ontologicamente dal suoi realizzatori fisici (pur nel senso debole che è richiesto lino stato fisico che realizzi M*, non necessariamente uno stato di tipo P*), sembra ragionevole sostenere che non è possibile che M causi M* senza nello stesso tempo causare P*. Se M è una causa autentica della mia intenzione di prendere un'aspirina, deve determinare in qualche modo lo stato fisico sottostante a tale intenzione. Dunque, affinché M causi M*, è necessario che M causi P*: la causalità mentale richiede la «causalità verso il basso» *(downward causation,* cfr. Fig. *b)*). Ciò, d'altra parte, è implicito nella possibilità, comunemente ammessa, che gli stati mentali causino stati fisici, come quando diciamo che una fitta dolorosa al gomito ci fa emettere un gemito.
La causalità verso il basso comporta però la spiacevole conseguenza della *sovradeterminazione causale*: P* risulta infatti causato tanto da P quanto da M. La sovradeterminazione causale potrebbe in qualche caso essere tollerabile ma, poiché l'argomento esposto ha validità generale, qui saremmo costretti ad accettare la poco attraente conclusione che c'è sovradeterminazione causale *ogni volta* che sono in gioco cause mentali. Si potrebbe evitare la sovradeterminazione causale invocando una priorità di P rispetto a M: P è «causa prima» di P* perché, ripercorrendo all'indietro la catena causale che ha condotto all'occorrenza di P*, sembra logico chiamare in causa un evento appartenente al medesimo ordine di spiegazione, cioè un evento cerebrale, piuttosto che uno mentale. In altri termini, non vi è motivo di fare ricorso a cause di natura estranea al fenomeno da spiegare, e specificamente a cause non fisiche, quando sono disponibili delle cause fisiche (è questo un modo semplice di enunciare il cosiddetto *principio di chiusura causale del mondo fisico*, che è comunemente considerato costitutivo del fisicalismo). Inoltre, almeno a detta di alcuni, invocare l'esistenza di una causa primaria fisica è non solo plausibile bensì necessario perché, in caso contrario (cioè senza la mediazione dello stato cerebrale), si dovrebbe ammettere la possibilità di relazioni causali «telecinetiche», cioè di azione causale a distanza. Necessario o soltanto ragionevole che sia, questo modo di evitare la sovradeterminazione causale ci riporta tuttavia a supporre che gli stati mentali, in questo caso M, siano causalmente inerti. Ricadiamo così nell'epifenomenismo cioè nella negazione di quella che era l'ipotesi di partenza: l'efficacia causale degli stati mentali. Sembra quindi che concepire la relazione tra mentale e fisico in termini di sopravvenienza ci ponga in ogni caso di fronte al problema di escludere una causa di troppo, vuoi per evitare la sovradeterminazione causale, vuoi perché non c'è bisogno di invocare una causa non fisica quando ce n'è già una fisica (sufficiente). Da cui il nome di «argomento dell'esclusione causale» (*causal/explanatory exclusion*).

Riassumendo, secondo Kim non è possibile conciliare la tesi dell'efficacia causale degli stati mentali con la sopravvenienza o altre equivalenti tesi antiriduzionistiche, se non incorrendo nell'epifenomenismo da un lato o nella sovradeterminazione causale dall'altro. Come sintetizza efficacemente Di Francesco, «se riconosciamo che il mentale sopravviene sul fisico, allora ad ogni supposta causa mentale sarà possibile associare una causa fisica, ma a questo punto le cause mentali non ci servono: [... 1 o si riducono a cause fisiche o non sono vere cause».

L' argomento dell'esclusione causale dimostra, secondo Kim, l'inevitabilità di una qualche forma di riduzione del mentale al fisico. Egli difende infatti una particolare versione di riduzionismo. Più in generale, Kim sostiene che il tentativo di spiegare la natura degli stati mentali sulla base della nozione di sopravvenienza è inevitabilmente una pseudosoluzione del problema mente-corpo, perché la sopravvenienza non è una *teoria* della relazione tra mente e mondo fisico. Essa è infatti compatibile con diverse teorie del rapporto tra mente e cervello, tutte insoddisfacenti. Tra queste Kini menziona l'epifenomenismo, in base al quale il mentale è distinto dal fisico ma è privo di efficacia causale e, soprattutto, *l'emergentismo,* secondo cui (semplificando un po') le proprietà mentali sono proprietà che un sistema fisico può manifestare soltanto a partire da una certa complessità di organizzazione, e non spiegabili né predicibili sulla base della sola struttura fisica. Ciò che in particolare sembra accomunare la sopravvenienza e l'emergenza è che il legame tra le proprietà di alto livello e quelle di basso livello rimane inspiegato. In questo senso la sopravvenienza sarebbe un'ipotesi dallo scarso o nullo valore esplicativo.

Il paradosso della causalità mentale

Come si potrebbe neutralizzare l'argomento dell'esclusione causale? Cominciamo con l'osservare che la difficoltà che esso mette in luce, lungi dall'essere «tecnica» o marginale, investe il problema centrale della metafisica della mente, il cuore stesso della questione mente-corpo: come possono gli stati mentali essere cause di stati fisici? Il problema della *causalità mentale* può essere presentato in modo semplice nella forma del seguente paradosso:

PARADOSSO DELLA CAUSALITÀ MENTALE

è impossibile che le tre seguenti affermazioni del senso comune siano simultaneamente vere:

1. Gli stati mentali sono distinti dagli stati fisici
2. Gli stati mentali sono cause di stati fisici (oltre che di altri stati mentali)
3. Solo gli stati fisici possono causare alcunché

A quale tesi rinunciare per venire a capo del paradosso? La candidata migliore sembra essere la prima. La tesi 2, infatti, è quella meno negoziabile, non soltanto per rivendicare il senso comune (non vogliamo smettere di dire che, per esempio, l'intenzione di farsi passare il mal di testa ha causato che il tale prendesse un'aspirina), ma anche allo scopo di difendere quella caratterizzazione degli stati mentali come mediatori tra input sensoriali e output comportamentali che è moneta corrente in scienza cognitiva. E sarebbe fonte di grave imbarazzo anche l'abbandono della tesi 3, nota come *principio della chiusura causale del mondo fisico* e comunemente considerata costitutiva del fisicalismo. Essa è esprimibile in modo più rigoroso nella forma «Ad ogni istante in cui uno stato fisico ha una causa, esso ha una causa fisica sufficiente». Ora, il modo più ovvio di negare la tesi 1 consiste nello sposare la teoria dell'identità di tipo: gli stati mentali risulterebbero identici agli stati cerebrali (strategia riduzionistica). Tuttavia, per i motivi che abbiamo visto, questa soluzione non è gradita al funzionalista, che ha cercato, tramite la teoria dell'identità di occorrenza e la nozione di sopravvenienza, un modo di negare la tesi i senza pagare, per così dire, un conto troppo salato - senza rinunciare alla realizzabilità multipla. L'argomento dell'esclusione causale colpisce, se conclusivo, precisamente questo tipo di strategia.

Il genere di dualismo che la teoria dell'identità di occorrenza o la nozione di sopravvenienza lasciano intravvedere può essere etichettato come «dualismo delle proprietà» o, da un punto di vista un po' diverso, come «dualismo epistemologico». Il *dualismo delle proprietà* consiste nella tesi secondo cui sebbene nel mondo ci siano esclusivamente enti (sostanze) particolari di natura fisica, questi sono dotati, oltre che di proprietà fisiche, anche di proprietà non fisiche. Le proprietà mentali sarebbero un caso di questo tipo. Il *dualismo epistemologico* consiste invece nella tesi secondo cui l'apparato concettuale che usiamo per caratterizzare i fenomeni mentali è irriducibile a quello neurologico. I due domini concettuali sono troppo diversi perché ci possano essere leggi-ponte o correlazioni empiriche evidenti tra stati mentali e stati fisici.
Per come è stata usata qui la nozione di proprietà, in modo ontologicamente non impegnativo, le due forme di dualismo non sono ben discriminabili: non ci siamo curati di discutere, per esempio, se e in che senso il *concetto* di credenza sia una «cosa» diversa dalla proprietà di essere una credenza. Preferisco tuttavia usare l'espressione «dualismo epistemologico», che meglio sottolinea come il genere di dualismo implicato dal funzionalismo riguardi il piano della spiegazione, più che il piano ontologico.
Il funzionalismo, interpretato non riduzionisticamente, implica dunque il dualismo epistemologico. Per i filosofi che si professano materialisti non riduzionisti il problema mente-corpo è divenuto il problema di venire a patti con questo tipo di dualismo. Quasi tutti respingono il cartesiano dualismo delle sostanze, e in questo senso il monismo ontologico, più che una tesi da dimostrare, è considerato un punto di partenza, un dato acquisito. Ma il dualismo epistemologico comporta nuovamente, come abbiamo visto, una sfida difficile: si tratta di una posizione coerente col fisicalismo? Le intuizioni sono diverse. Chi ritiene di sì insiste sul fatto che all'interno delle scienze della natura ci sono moltissime proprietà non riducibili a proprietà fisiche, senza che ciò tuttavia ne metta in discussione la loro dipendenza ontologica da proprietà fisiche. Per esempio, la *fitness*, la proprietà biologica che un organismo ha di essere bene adattato all'ambiente, non è riducibile a una proprietà fisica ma è certamente dipendente da un complesso di proprietà fisiche, *sopravviene* a una proprietà fisica (Chalmers). Chi difende questa esistenza di prospettiva ritiene che il tipo di dualismo implicato dalle proprietà funzionali non riducibili non sia tale da impensierire un materialista, perché ciò che conta è la dipendenza ontologica (diversamente dal caso delle proprietà *non* sopravvenienti, quale è, secondo Chalmers, la coscienza fenomenica; per Chalmers, la coscienza viola davvero il fisicalismo). L'intuizione di chi, invece, ritiene che il dualismo epistemologico sia incompatibile col fisicalismo è che quest'ultimo comporta che ogni proprietà vada in ultima analisi spiegata in un vocabolario fisico. Per esempio, Bickle afferma che il cosiddetto fisicalismo non riduzionistico non è affatto fisicalismo, in quanto nega che una scienza fisica matura possa spiegare gli aspetti essenziali del mentale. In questa prospettiva non può esserci un autentico monismo ontologico in concomitanza col dualismo epistemologico. Non si può tranquillamente sostenere che le proprietà mentali dipendono ontologicamente dalle proprietà fisiche se non si è in grado di giustificare, invocando leggi opportune, questa dipendenza, se non si riesce *spiegare* il mentale col fisico. Il punto è, dunque, che non c'è accordo su che cosa si debba esattamente intendere per «fisicalismo».

In conclusione, una qualche forma di dualismo sembra essere l'esito tipico del funzionalismo applicato al problema mente-corpo, ed anzi è probabilmente l'esito inevitabile di qualunque posizione che cerchi di coniugare il monismo[7] ontologico con la peculiarità del mentale. L'atteggiamento metafisico oggi più diffuso in filosofia della mente consiste quindi nel mettere insieme monisino ontologico e dualismo epistemologico. Ironicamente, siamo così tornati a una situazione molto cartesiana, con le proprietà (o i concetti) al posto delle sostanze: da un lato abbiamo le proprietà mentali, dall'altro le proprietà fisiche, e non sappiamo bene se e come correlarle. Secondo alcuni ciò dimostra che il problema niente-corpo è ancora attuale e lontano dall'avere trovato una soluzione; secondo altri questa situazione è del tutto soddisfacente e non solleva un genuino problema filosofico.

[7] Al contrario di dualismo, si chiama "monismo" (dal greco "*mònos*", "solo", cfr. "monaco", "monoteismo", "monarchia" ecc.) ogni teoria che ammette un solo tipo di "realtà" (o sostanza) a cui tutte le altre sono riducibili (tipicamente: la dottrina di Spinoza).