

**Parte II**

**Sopravvenienza ed emergenza in una prospettiva fiscalistica**

### 3. Il concetto di sopravvenienza dall'etica alla filosofia della mente

#### 3.1 *Fisicalismo e riduzionismo: impossibile separarli?*

Nella parte I abbiamo introdotto il problema mente-corpo e le principali soluzioni che sono state proposte dalla filosofia contemporanea. Abbiamo anche visto alcuni tra i tanti limiti di tali proposte, e in particolare: la tendenza del *dualismo interazionista* ad allontanarsi non solo dall'epistemologia della biologia, ma addirittura dall'indagine puramente razionale, a favore della fede nel soprannaturale; il fraintendimento di ciò che in una teoria della mente può essere considerato esplicativo, da parte del paradigma opposto al dualismo, il *materialismo eliminativista* (almeno nella sua versione più caratteristica, cioè l'eliminativismo "forte" verso la psicologia del senso comune); l'impossibilità del confronto costruttivo con le neuroscienze per l'approccio "soggettivistico-personologico". La posizione di Karl Popper è stata esaminata con più attenzione e con essa abbiamo incontrato una prima formulazione del paradigma emergentista, molto vicina all'impegno sostanzialista del dualismo cartesiano - tanto da opporsi appassionatamente all'orientamento "naturalistico-fisicalista" oggi dominante in filosofia della mente.

In questa seconda parte (capitoli 3 e 4) ripartiremo dalla posizione epistemologica del paradigma dominante della scienza cognitiva, il *funzionalismo*, per compiere un ulteriore avvicinamento all'epistemologia della biologia - rimanendo comunque nell'ambito dell'approccio "naturalistico-fisicalista" (che non abbandoneremo più). Per varie ragioni, in filosofia della mente, muoversi nell'area epistemologica *intermedia* tra dualismo ed eliminativismo comporta una certa

“instabilità epistemologica”, alla quale dovremo far fronte prestando particolare attenzione all’uso delle parole e dei concetti chiave<sup>1</sup>.

Va notato che il funzionalismo, il paradigma dominante, *non* è un indirizzo filosofico particolarmente vicino all’epistemologia della (neuro)biologia. Per esempio, benché la maggior parte dei suoi esponenti condivide senza remore l’impegno ontologico fisicalista, *dal punto di vista logico* la natura funzionale della mente non lo richiede: neuroni, diodi e *quanta* di energia ectoplasmatica potrebbero essere organizzati in modo tale da avere stati mentali. Inoltre, lacuna molto grave, il funzionalismo non ha mai adottato quella prospettiva evuzionistica che è invece tipica della terza cultura, cioè non ha inserito il problema della mente nel contesto del darwinismo. Insomma, con il funzionalismo ci si trova ancora epistemologicamente lontani dalla biologia - e in particolare dalla biologia evuzionistica.

Cominciamo ora ad esaminare quali assunzioni comporta l’avvicinamento epistemologico alla (neuro)biologia. L’assunzione principale è ovviamente l’ontologia fisicalista. All’inizio del capitolo 1 avevamo osservato che per minore distanza epistemologica dalla biologia non si deve tanto intendere un avvicinamento all’atteggiamento riduzionistico, quanto piuttosto una maggiore compatibilità con le conoscenze teorico-sperimentali della neurobiologia. Ora, però, adottando l’ontologia fisicalista, è necessario precisare anche se (e, se sì, in quale misura) tale ontologia comporta un atteggiamento riduzionistico. Infatti, mentre in un contesto epistemologicamente lontano dalla biologia *riduzione a e compatibilità con* la neurobiologia sono ben distinte, nel nuovo ambito di discorso tale distinzione tende a sfumarsi. Inoltre, il rapporto tra fisicalismo e riduzionismo è per la nostra ricerca un problema fondamentale, perché se si stabilisse che il fisicalismo *comporta* il riduzionismo, allora l’emergentismo, spesso considerato addirittura il caso

---

<sup>1</sup>Perché avvicinarsi alla biologia? Ovvero: perché complicarsi (filosoficamente) la vita rinunciando alle posizioni epistemologicamente stabili? A questa domanda si può rispondere a molti livelli diversi. La risposta più “metafisica” potrebbe essere che, secondo una parte importante della filosofia del Novecento, *la verità umana è biologica*. In altre parole: gli esseri umani, più di qualsiasi altra qualifica, sono esseri biologici. Secondo questa tesi, avvicinarsi all’epistemologia della biologia, per la filosofia della mente, significa avvicinarsi alla *verità* della mente. E’ superfluo sottolineare che non si può *dimostrare* in alcun modo una tesi metafisica; si potrebbe *argomentare* a suo favore, ma la sede non è quella opportuna. Quindi questa risposta è destinata a costituire, al massimo, lo “sfondo metafisico” degli spostamenti filosofici verso l’epistemologia della biologia e, in particolare, quello degli spostamenti della filosofia della mente verso l’epistemologia della neurobiologia (la quale, è bene precisarlo, non è necessariamente l’epistemologia *dei neurobiologi*).

paradigmatico di fisicalismo non-riduzionista, sarebbe contraddittorio (sosterrebbe due tesi di cui una *comporta* la negazione dell'altra). Ma, sebbene alcuni autori non abbiano dubbi in proposito, prima di “gettare la spugna”, occorre esaminare che cosa si intende esattamente per «fisicalismo» e che cosa per «riduzionismo».

Scopo di questo paragrafo sarà illustrare *alcuni* dei significati di queste due parole, evidenziando così il fatto che non ha senso sostenere, genericamente, che *il* fisicalismo comporta *il* riduzionismo: è invece corretto dire che alcune *forme* di fisicalismo comportano certe *forme* di riduzionismo (o addirittura coincidono con esse), mentre altre no.

Iniziamo a esaminare l'ipotesi fisicalista<sup>2</sup>. Di essa abbiamo già incontrato l'interpretazione di Popper, secondo cui: (a) *il mondo è un sistema fisico chiuso*. Questa formulazione non è priva di ambiguità perché per “mondo fisico” si può intendere sia il “mondo descrivibile mediante le leggi della fisica”, sia il “mondo non-soprannaturale”.

b) Formulazione “negativa” del fisicalismo: *le leggi fisiche non possono essere violate*. Questa formulazione ci permette di eliminare dalle teorie della mente ogni sorta di sostanza non-fisica: spiriti, entelechie, ecc. Tuttavia, essa è molto debole e la maggior parte degli autori appartenenti all'indirizzo “naturalistico-fisicalista”, di solito, la considera ovvia e scontata.

c) Fisicalismo come “chiusura *causale* del mondo fisico”. Molti autori ritengono che secondo la scienza moderna il mondo fisico - fisicamente descrivibile - è causalmente chiuso [cfr. Putnam 1981, 83], cioè i fenomeni fisici possono interagire *causalmente* solo tra loro. Si potrebbe allora adottare questa condizione come interpretazione del fisicalismo [cfr. Horgan 1993]. Tuttavia, anche il concetto di causalità è elusivo e la sua precisazione è forse più complessa di quella del fisicalismo stesso<sup>3</sup>.

---

<sup>2</sup>Salvo avviso contrario, useremo i termini «materialismo» e «fisicalismo» come sinonimi, intendendo con entrambi la versione “auto-trascesasi” del monismo materialista [Popper 1977, §3]. Tuttavia, per evitare inutile confusione, si cercherà di evitare il termine «materialismo» e, laddove gli autori esplicitamente ne dichiarino la sinonimia, si parlerà di «fisicalismo».

<sup>3</sup>Mario Bunge, per esempio, sostiene che «una cosa o una proprietà di una cosa non potrebbero essere delle cause: soltanto i cambiamenti possono avere un'efficacia causale» [Bunge 1971, 134]; e a questo proposito specifica ben sei condizioni che devono essere soddisfatte affinché la relazione tra cambiamenti di due o più proprietà possa essere chiamata *causale* [Bunge 1971, 135-6].

d) Fisicalismo come “dipendenza dai fatti fisici”. Secondo molti studiosi, i fisicalisti sostengono la tesi secondo cui i fatti fisici, in qualche modo, *determinano* tutti i fatti del mondo. Ci sarebbe dunque un’asimmetria tra ciò che è fisico (in senso stretto, cioè *descrivibile fisicamente*) e ciò che non lo è. Non solo: alcuni autori associano il fisicalismo al tentativo di sviluppare dei veri e propri “principi di determinazione fisica” che esplicitino e spieghino l’asimmetria tra fatti fisici e fatti non-fisici<sup>4</sup>.

e) Fisicalismo come teoria dell’identità mente-cervello. Come se non bastassero tutte le formulazioni precedenti, a complicare ulteriormente le cose è il fatto che in filosofia della mente il fisicalismo si trova spesso associato alla teoria dell’identità mente-cervello. Dennett, per esempio, scrive:

l’affermazione che «le menti non sono altro che i cervelli» è un modo molto informale per esprimere una versione di *fisicalismo*, ovvero della *teoria dell’identità* della mente (così chiamata perché *identifica* gli eventi mentali con eventi *fisici* nel cervello) [Dennett 1978a, 18].

Poiché esistono due versioni della teoria dell’identità, cioè l’identità dei *tipi* e l’identità delle *occorrenze*, si possono trovare le espressioni: “fisicalismo dei tipi” e “fisicalismo delle occorrenze” [Fodor 1981, 23], e ancora: “materialismo del tipo” e “materialismo dell’occorrenza” [Di Francesco 1996, 91-3].

Prima di qualsiasi commento, volgiamo l’attenzione al riduzionismo. Per presentare le interpretazioni più frequenti di questo concetto, seguiremo parzialmente John Searle, il filosofo della mente che ha introdotto la distinzione tra proprietà «emergenti 1» ed «emergenti 2», il quale elenca cinque forme di riduzione (alle quali, è ovvio, corrispondono altrettante forme di riduzionismo):

1) *Riduzione ontologica*. E’ la forma più importante di riduzione, ma anche la più vaga: attraverso di essa è possibile mostrare che «oggetti di determinati tipi sono nient’altro che oggetti di altri tipi» [Searle 1992, 128]. Per esempio, dal punto di vista della riduzione ontologica, tutti gli oggetti materiali macroscopici non sono altro che collezioni di molecole.

---

<sup>4</sup>Cfr. per esempio: Hellman - Thompson [1975], “Physicalism: Ontology, Determination, and Reduction”, in «Journal of Philosophy», n.72 (1975), pp.551-64; citati in Kim [1993b, 341].

2) *Riduzione delle proprietà*. Spesso conseguenza di quella tra teorie (3), è anche questa una forma di riduzione ontologica. L'esempio canonico è la riduzione della temperatura, che è una proprietà dei sistemi macroscopici (per esempio i gas), all'energia cinetica media molecolare, che è una proprietà delle componenti microscopiche dei sistemi [cfr. equazione (c) alla nota n.5 in questa pagina].

3) *Riduzione nomologica o tra teorie*. Si tratta della forma di riduzione di cui si dibatte maggiormente, anche e soprattutto in riferimento all'opera di Ernest Nagel, *The Structure of Science* (1961), nella quale viene discussa dettagliatamente [Nagel 1961, cap.11]. La riduzione tra teorie è collegata strettamente a quella delle proprietà; con la differenza, però, che in questo caso non si tratta di proprietà ma di *leggi o equazioni*. L'esempio ricorrente di questa riduzione è quello della termodinamica classica alla teoria cinetica dei gas<sup>5</sup>.

4) *Riduzione logica (o linguistica)*. Secondo Searle, questo tipo di riduzione avviene tra le parole e gli enunciati che si riferiscono ad entità tra loro riducibili ontologicamente. Per esempio, gli «*enunciati relativi ai numeri [...] possono essere tradotti in - e quindi ridotti a - enunciati relativi a insiemi*» [Searle 1992, 129]. Quando il termine «fisicalismo» fu introdotto per la prima volta nel dibattito filosofico (dai neopositivisti), esso indicava una forma di riduzionismo associata alla riduzione logico-linguistica di tutti gli enunciati dotati di senso agli enunciati della fisica.

---

<sup>5</sup>Presentiamo qui, in forma semplificata, tale riduzione nomologica [Nagel 1961, 353-4]. L'idea intuitiva è che le equazioni della termodinamica classica siano deducibili, tramite certe "leggi-ponte", da quelle della teoria cinetica dei gas. L'equazione più importante della termodinamica classica è la nota equazione di stato dei gas (*empiricamente* ricavata):  $PV=nRT$ , dove  $P$  è la pressione,  $V$  il volume,  $n$  il numero di moli e  $T$  la temperatura del gas e  $R$  è la costante universale dei gas. Alcuni gas (idrogeno, elio, argon, azoto) seguono questa legge con eccellente approssimazione; altri (per esempio l'anidride carbonica) si avvicinano ad essa solo a temperature elevate; per definizione si chiama "gas perfetto" il gas ideale che segue *esattamente* l'equazione di stato. L'equazione di stato può essere scritta nella seguente forma del tutto equivalente alla prima: (a)  $PV=NKT$ , dove  $N$  è il numero di molecole del gas e  $K$  è la costante di Boltzmann. *Indipendentemente* dalla termodinamica classica, lo studio microscopico dei gas ha condotto all'*interpretazione microscopica* dei parametri di stato dei gas (pressione, volume). Il risultato più importante di questa interpretazione è l'equazione: (b)  $PV=2NE/3$ , dove  $E$  è l'energia cinetica *media* delle molecole del gas. Ora, date le due equazioni (a), fornita dalla termodinamica classica, e (b), fornita dalla teoria cinetica dei gas, risultano equivalenti le seguenti affermazioni:

- Dal confronto di (a) e (b) si ottiene la seguente equazione, valida per il gas perfetto: (c)  $E=3KT/2$ ;
- L'equazione (a) può *ridursi nomologicamente* a (b), cioè (a) può essere dedotta da (b), tramite la "legge-ponte" (c).

5) *Riduzione causale*. Secondo Searle si ha riduzione causale quando «*l'esistenza, e a fortiori i poteri causali, dell'entità soggetta a riduzione possono essere interamente spiegati in base ai poteri causali dei fenomeni a cui essa viene ridotta*» [Searle 1992, 129]. Tuttavia, se si preferisce evitare la difficoltà concettuale implicata dall'uso dell'espressione «poteri causali», si può adottare la seguente formulazione: (6) si ha questo tipo di riduzione quando «*qualsiasi proprietà di un sistema complesso (l'intero) è completamente determinata dalla struttura del sistema, cioè dal numero, dalle caratteristiche e dalle mutue relazioni delle parti costituenti*» [Amsterdamski 1981, 64]. Naturalmente, così interpretato, il riduzionismo coincide con il *determinismo* (o almeno con un'accezione di esso).

Applicando ciascuna di queste interpretazioni del riduzionismo al problema mente-corpo si ottengono le seguenti affermazioni:

- 1) (Riduzione ontologica) La mente non è altro che il cervello;
- 2) (Riduzione delle proprietà) Le proprietà mentali non sono altro che proprietà cerebrali;
- 3) (Riduzione nomologica o tra teorie) Esistono leggi di correlazione tra “leggi psichiche” e “leggi neurofisiologiche”;
- 4) (Riduzione logica o linguistica) Gli enunciati relativi alla mente sono traducibili in enunciati relativi al cervello;
- 5) (Riduzione causale) Le «*proprietà della mente sono causate da processi neurobiologici*» [Searle 1992, 130];
- 6) (Determinismo) Le proprietà mentali del sistema nervoso sono completamente determinate dalla sua struttura, cioè dal numero, dalle caratteristiche e dalle mutue relazioni delle parti costituenti.

Come vedremo alla fine del capitolo, oltre ad essere associabile a molte tesi specifiche differenti, il riduzionismo può anche essere interpretato come un atteggiamento la cui assunzione è *questione di grado* e non è una scelta del tipo “tutto o niente”.

Per ora si può concludere, alla luce delle due serie di accezioni dei termini «fysicalismo» e «riduzionismo» presentate in questo paragrafo, che qualunque presunta relazione tra l'ipotesi fysicalistica e l'atteggiamento riduzionistico andrebbe discussa precisando *a quali forme di essi* ci si riferisce. Per esempio, è ovvio che se

intendiamo, per fisicalismo, la posizione dei teorici dell'identità mente-cervello (come fa Dennett) e, per riduzionismo, la posizione di chi difende la riduzione *ontologica* della mente al cervello, allora le due posizioni *coincidono*. Analogamente, se sosteniamo, come fisicalisti, la tesi secondo cui i fatti fisici determinano tutti i fatti del mondo e, come riduzionisti, quella del determinismo, nuovamente le due tesi - in pratica - *coincidono*. Tuttavia, non tutte le possibili combinazioni delle accezioni dei due termini in questione sono così banali.

Un caso molto interessante è quello del *monismo anomalo* del filosofo Donald Davidson, che unisce il fisicalismo nell'accezione (e) all'esplicito rifiuto del riduzionismo nell'accezione (3)<sup>6</sup>.

### 3.2 *Il monismo anomalo di Davidson*

Il monismo anomalo, presentato da Davidson nel suo citatissimo saggio "Mental Events" (1970), è una delle versioni più controverse dell'identità mente-cervello nella forma che viene chiamata "identità delle *occorrenze*"<sup>7</sup>.

La possibilità stessa di un'identità tra eventi mentali e cerebrali, dice Davidson, è offuscata da quelle formulazioni dell'identità mente-cervello che tendono «a oscurare la distinzione fra eventi particolari e generi di evento» [Davidson 1970, 292]. Per esempio, secondo Davidson, quando il teorico dell'identità Smart paragona l'identità mente-cervello a quella "fulmine-scarica elettrica", precisando che

---

<sup>6</sup>Si può equivalentemente dire che il monismo anomalo è una forma di riduzionismo *ontologico* senza riduzione *nomologica*. È facile convincersi del fatto che, in qualche modo, questa posizione comporta il rifiuto del riduzionismo *delle proprietà*. Per questa ragione, alcuni autori considerano il monismo anomalo una forma di *dualismo delle proprietà*, orientamento riduzionista nei confronti della sostanza (accezione 1) e anti-riduzionista nei confronti delle proprietà (accezione 2). Tuttavia, benché la differenza tra monismo anomalo e dualismo delle proprietà assomigli un po' a quella tra il bicchiere "mezzo pieno" e il bicchiere "mezzo vuoto", vi sono buone ragioni per mettere in secondo piano la questione del dualismo delle proprietà. Il punto cruciale di tutta la discussione, infatti, è quale posizione si ha nei confronti del riduzionismo *nomologico*. Davidson e Kim, entrambi dualisti delle proprietà, si trovano contrapposti circa tale problema.

<sup>7</sup>L'identità delle *occorrenze* è un po' più debole dell'identità dei *tipi*, che la comprende, perché si limita ad affermare l'identità tra singoli eventi mentali e singoli eventi cerebrali. L'identità dei *tipi*, invece, richiede: a) «che ogni specifica occorrenza di uno stato mentale sia identica all'occorrenza di uno stato cerebrale» [Di Francesco 1996, 92] (identità delle occorrenze); ma anche: b) «che ogni ulteriore occorrenza di quel tipo di stato mentale sia identica a un'occorrenza del tipo di stato cerebrale originariamente associato» [*ibid.*].



non ci sono due cose: un fulmine e una scarica elettrica. C'è una cosa sola, un fulmine, che si descrive scientificamente come una scarica elettrica verso terra da una nube di molecole d'acqua ionizzate<sup>8</sup>

- ciò andrebbe inteso come un caso di identità delle occorrenze, cioè: *«per ogni fulmine, esiste una scarica elettrica verso terra da una nube di molecole d'acqua ionizzate tale che il fulmine è identico a essa»* [Davidson 1970, 292]. Esclusivamente in questa formulazione abbiamo *«un'onesta ontologia di eventi individuali, e possiamo intendere l'identità in senso letterale»* [ibid.].

Il *monismo anomalo* è la posizione secondo cui tra eventi mentali ed eventi cerebrali vi è identità delle occorrenze (*monismo*) e tuttavia non vi sono leggi di correlazione psico-fisiche rigorose (*anomalia*).

Questa tesi è *fisicalista*, addirittura nella versione del fisicalismo come “teoria dell'identità”, e nello stesso tempo *non-riduzionista*; *«il mentale è nomologicamente irriducibile: possono esserci asserzioni vere che mettono in relazione il mentale e il fisico, asserzioni che hanno la forma logica di una legge; ma non sono di carattere nomico»* [Davidson 1970, 295-6]. Qual è la differenza tra asserti semplicemente *legisimili* e leggi o asserti *nomologici*?

Secondo Davidson, gli *«asserti nomologici mettono insieme predicati dei quali sappiamo a priori che sono fatti l'uno per l'altro: cioè lo sappiamo indipendentemente dal fatto di sapere se ci sono prove a sostegno di una connessione fra essi»* [Davidson 1970, 298-9]. L'esempio di Davidson è il seguente:

«Tutti gli smeraldi sono verdi» ha carattere nomico in quanto tutti i suoi esempi lo confermano, ma «Tutti gli smeraldi sono blerdi» non lo è, in quanto «blerde» significa «osservato prima dell'istante *t* e verde, altrimenti blu», e, se tutte le nostre osservazioni fossero compiute prima di *t* e rivelassero uniformemente smeraldi verdi, questa non sarebbe una ragione per aspettarci che altri smeraldi siano blu. Nelson Goodman ha congetturato che questo dimostri come certi predicati, per esempio «blerde», non sono adatti alle leggi [...]. A me sembra però che il carattere anomalo di «Tutti gli smeraldi sono blerdi» dimostri soltanto che i predicati «è uno smeraldo» e «è blerde» non sono reciprocamente adatti [...]. «Blu», «rosso» e «verde» sono fatti per gli smeraldi, gli zaffiri e le rose; «verlù», «blerde» e «verosso» sono fatti per gli zaffaldi, gli smeriri e le smerose [Davidson 1970, 298-9].

---

<sup>8</sup>Smart J. J. C. [1959], “Sensations and Brain Processes”, in Chappell V. C. (a cura di), *The Philosophy of Mind*, Englewood Cliffs, New Jersey 1962, pp.163-5; citato in Davidson [1970, 291-2].

L'analogia a questo punto è prevedibile: i predicati fisici e quelli mentali non sono reciprocamente adatti e la formulazione di una legge che li legasse non sarebbe nomologica; dal punto di vista del carattere nomico, infatti, «*gli asserti psicofisici assomigliano più a "Tutti gli smeraldi sono blerdi" che a "Tutti gli smeraldi sono verdi"»* [Davidson 1970, 299].

L'idea, che in filosofia della mente non è una sorpresa, è dunque quella di una *differenza di categoria* tra il mentale e il fisico. Da tale differenza categoriale all'impossibilità di leggi psicofisiche rigorose, poi, il passo è breve. La vera novità del monismo anomalo, dice Davidson, sta nel fatto che la negazione del riduzionismo nomologico *contribuisce* a stabilire l'identità (delle occorrenze) tra il mentale e il fisico. Vediamo come.

L'argomentazione parte dall'ipotesi che siano imprescindibili i seguenti principi:

- Principio di interazione causale: «*almeno alcuni eventi mentali interagiscono causalmente con eventi fisici*» [Davidson 1970, 286];
- Principio del carattere nomologico della causalità: «*ciascun asserto causale singolare vero è suffragato da una legge rigorosa che collega gli eventi [...] menzionati in qualità di causa e di effetto*» [Davidson 1970, 305].

Si consideri ora un evento mentale che, per il principio di interazione causale, interagisce causalmente con un evento fisico. Allora, per il secondo principio, deve esistere una legge rigorosa che collega i due eventi; tuttavia, se rifiutiamo il riduzionismo nomologico, non può esistere una legge psicofisica rigorosa tra eventi mentali e eventi fisici: quindi, se legge deve esistere, essa può essere solo fisica. Ciò significa che *entrambi* gli eventi sono collegati da una legge fisica. Ma se anche l'evento mentale ricade sotto una legge fisica, allora possiede una *descrizione fisica*; quindi, per definizione, è un evento fisico. «*Dunque ogni evento mentale che sia causalmente correlato a un evento fisico è un evento fisico*» [Davidson 1970, 306].

Questa argomentazione costituisce il nucleo di "Mental Events". Tuttavia, essa non ha avuto molta risonanza. Ciò che ha reso famoso questo saggio è stata invece la seguente, importantissima precisazione:

Sebbene la posizione da me descritta neghi l'esistenza di leggi psicofisiche, è compatibile con l'idea che le caratteristiche mentali sono in qualche senso dipendenti da, o che sopravvengono a caratteristiche fisiche. Si può intendere tale sopravvenienza nel senso che non possono esserci due eventi simili in tutti gli aspetti fisici, ma diversi per qualche aspetto mentale; oppure che un oggetto non può mutare per qualche aspetto mentale senza mutare per qualche aspetto fisico. Tale dipendenza o sopravvenienza non implica una riducibilità mediante leggi o definizioni [Davidson 1970, 293-4].

Con queste parole, Davidson realizzava un "trapianto" concettuale oggi noto come l'«appropriazione materialistica della sopravvenienza» [Horgan 1993, 563]. In questo passo, infatti, egli applica alla relazione mente-cervello il concetto di *soppravvenienza*, mutuandolo direttamente dall'etica, dove era stato elaborato per formalizzare la relazione tra proprietà morali e proprietà fisiche.

L'appropriazione del concetto di sopravvenienza da parte dei filosofi della mente ha un ruolo di primo piano nella storia contemporanea del problema mente-corpo. Tra il 1970 e i primi anni '90, infatti, molti filosofi (soprattutto di tradizione analitica) hanno cercato di utilizzare il concetto di sopravvenienza per *dare una formalizzazione logica al rapporto tra la mente e il cervello*. Nel fare questo alcuni hanno sperato, come Davidson, di poter adottare il fisicalismo senza abbracciare il riduzionismo; tuttavia, il filosofo che più di chiunque altro ha studiato il concetto di sopravvenienza, Jaegwon Kim, si è battuto con passione *contro* la possibilità di un fisicalismo non-riduzionista (in particolare, nella accezione del fisicalismo come "dipendenza dai fatti fisici" e in quella *nomologica* del riduzionismo).

Secondo Kim, le posizioni sostenibili relative al problema mente-corpo sono solo tre: dualismo (anti-fisicalista), *riduzionismo*, eliminativismo [Kim 1989, 267]. Se applichiamo questa prospettiva alla schematizzazione che abbiamo proposto all'inizio del primo capitolo, se ne deduce che tutte le posizioni "intermedie" rispetto al dualismo e all'eliminativismo sono compromesse con il riduzionismo o addirittura che sono incoerenti. Possiamo fare simili affermazioni anche a proposito del funzionalismo e del monismo anomalo?

Kim non ha dubbi: il monismo anomalo di Davidson è *virtualmente indistinguibile dall'eliminativismo* [Kim 1989, 270]. Il *funzionalismo* di Putnam, d'altra parte, comporta una sorta di riduzionismo "specie-specifico", che è *un riduzionismo in piena regola* [Kim 1989, 274].

Il problema comune a queste correnti, e a tutti gli orientamenti filosofici che tentano la strada del fisicalismo non-riduzionista, è il seguente. Dal punto di vista logico, la relazione tra mente e cervello deve essere contemporaneamente:

- 1) così *forte* (logicamente) da far sì che la mente *dipenda* dal cervello (fisicalismo come “dipendenza dai fatti fisici”);
- 2) così *debole* (logicamente) da far sì che la mente *non si riduca* al cervello (riduzione nomologica).

Secondo Kim, *conciliare queste due clausole non è possibile*. E, in particolare, non è possibile *neppure utilizzando il concetto di sopravvenienza*. Kim è convinto che *la mente sopravvenga sul cervello, ma è anche convinto che la forma di sopravvenienza implicata nella relazione mente-cervello sia troppo “forte” per poter parlare di non-riducibilità*. Per illustrare questa posizione, seguiremo da vicino Kim [1990], articolo in cui vengono esaminate varie relazioni logico-formali, tutte “candidate” al ruolo di *sopravvenienza*.

### 3.3 *Sopravvenienza, covarianza, dipendenza, non-riducibilità*

Il concetto di sopravvenienza, come abbiamo detto, è nato nel contesto della filosofia morale. Autori come George E. Moore e Richard M. Hare hanno combattuto le definizioni di proprietà morali date in termini non-morali (“fallacia naturalistica”) e hanno elaborato il concetto di sopravvenienza per rendere conto della relazione tra attributi etici e non. Moore è stato il primo, negli anni Venti, ad introdurre l’idea (non la parola), parlando di “valore intrinseco”: se una data cosa possiede un qualsiasi valore intrinseco in un certo grado, allora non solo necessariamente lo possiederà, in tutte le circostanze, nel medesimo grado, ma qualsiasi altra cosa *esattamente come quella* possiederà necessariamente lo stesso valore intrinseco, in tutte le circostanze e nel medesimo grado. Hare, in *The Language of Morals* (1952), ha ripreso quest’idea e l’ha associata al termine «sopravvenienza».

Secondo Kim, Moore e Hare hanno in mente un nesso di *covarianza necessaria* tra tipi di proprietà: *un cambiamento nelle proprietà di un tipo non può avvenire a meno che non sia accompagnato da un cambiamento nelle proprietà di un*

*altro tipo* [Kim 1990, 137]. Ma, a questa caratteristica fondamentale, il monismo anomalo di Davidson, applicando la sopravvenienza al problema mente-corpo, aggiunge altri due elementi: la dipendenza e la non-riducibilità<sup>9</sup>. In altre parole, sembra che la relazione di sopravvenienza tra proprietà debba soddisfare contemporaneamente le seguenti tre condizioni [Kim 1990, 140; Davidson 1970, 293-4]:

- *Covarianza*: le proprietà sopravvenienti covariano con le loro proprietà di base;
- *Dipendenza*: le proprietà sopravvenienti dipendono dalle loro proprietà di base;
- *Non-riducibilità*: le proprietà sopravvenienti sono non-riducibili alle loro proprietà di base.

Ovviamente, la covarianza è la condizione cruciale: *se non c'è covarianza, non c'è sopravvenienza*. Si tratta di vedere, a questo punto, se la covarianza può consentire sia la dipendenza, sia la non-riducibilità. Se sì, allora tale relazione deve essere considerata la forma di sopravvenienza di cui ha bisogno il fisicalismo non-riduzionista per formalizzare logicamente il rapporto mente-cervello.

Kim presenta due tipi principali di covarianza (a cui poi se ne aggiungeranno altri due). Entrambi fanno uso di alcuni concetti appartenenti alla filosofia analitica che qui verranno proposti in forma leggermente semplificata: il concetto di *mondo possibile* e il concetto di *indiscernibilità* (a questi aggiungeremo, parlando del quarto tipo di covarianza, il concetto di proprietà *qualitativa intrinseca*).

Il concetto di *mondo possibile* ha una storia lunga e complessa: dai logici medievali a Hugh Everett III, da Leibniz a Saul Kripke, esso è stato usato da autori eterogenei e per scopi spesso diversi. Ai nostri fini teorici, basterà considerare il mondo possibile un'entità logica astratta, *relativamente alla quale* è possibile interpretare espressioni linguistiche senza contraddirsi. Ai fini pratici, si tratta

---

<sup>9</sup>Kim nota giustamente che (a) la non-riducibilità (*nonreductiveness*), che Davidson attribuisce agli eventi mentali, non va confusa con (b) l'irriducibilità (*irreducibility*); infatti, mentre (a) *non implica* la riduzione, (b) *implica l'impossibilità* di riduzione. Tuttavia, dopo questa precisazione [Kim 1990, 139], Kim passa indifferentemente da (a) a (b) [Kim 1990, 140]. Nel 1993 Kim ha poi lasciato perdere del tutto la condizione della non riduzione (sia (a) che (b)), circoscrivendo le questioni relative alla sopravvenienza alle clausole di covarianza e dipendenza [Kim 1993a, 165]. Nel presente lavoro, tuttavia, noi seguiremo la raccomandazione di non confondere (a) e (b): pertanto, dicendo che una proprietà è *non-riducibile*, intenderemo che *non esiste* di fatto una sua riduzione; diremo invece che una proprietà è *irriducibile* quando *non può esistere* per principio una sua riduzione.

semplicemente di un mondo “pensabile”, cioè coerente dal punto di vista logico<sup>10</sup>. Per quanto riguarda l’altro concetto, basti dire che, considerato un insieme A di proprietà qualsiasi, l’*indiscernibilità* rispetto ad A (cioè la *A-indiscernibilità*) è l’impossibilità di distinguere qualcosa relativamente alle proprietà di A.

Sia M l’insieme delle proprietà *sopravvenienti* sull’insieme di proprietà F; Kim considera le relazioni seguenti come possibili forme di sopravvenienza [Kim 1990, 141]<sup>11</sup>:

1. *Covarianza debole*: per ogni mondo (logicamente) possibile contenente x e y, se x e y sono F-indiscernibili, allora sono M-indiscernibili;
2. *Covarianza forte*: per ogni x(w) e y(w’) appartenenti rispettivamente ai mondi (logicamente) possibili w e w’, se x(w) e y(w’) sono F-indiscernibili, allora sono M-indiscernibili.

Almeno uno di questi due tipi di covarianza può essere una forma di sopravvenienza se la sopravvenienza deve essere una relazione di *dipendenza*? E può esserlo se la sopravvenienza deve essere una relazione di *non-riducibilità*? Kim considera separatamente le due questioni.

Cominciamo con la *dipendenza*. A questo proposito, il problema della covarianza *debole* è di essere troppo debole. Infatti, poiché essa opera solo entro un singolo mondo possibile alla volta, qualsiasi cosa può accadere negli altri mondi possibili. Applicando la covarianza debole al rapporto mente-cervello, essa risulta per esempio compatibile con i seguenti tre mondi (logicamente) possibili, *anche se assumiamo per ipotesi che siano fisicamente-indiscernibili dal nostro* [Kim 1990, 143]:

- un mondo in cui non esistono menti;
- un mondo in cui gli esseri unicellulari hanno una mente, ma gli umani no;
- un mondo in cui ogni cosa ha le stesse identiche facoltà mentali.

---

<sup>10</sup>In realtà, vi è un’importante differenza tra mondi possibili *logicamente* e possibili *fisicamente*. I primi sono semplicemente *non-incoerenti*; i secondi, invece, sono logicamente possibili e *inoltre non contraddicono le leggi della fisica* (attualmente note). Benché per ogni forma di covarianza andrebbero considerate le due versioni, quella che fa uso del concetto di mondi logicamente possibili e quella che fa uso del concetto di mondi fisicamente possibili, noi ci limiteremo alla versione dei mondi *logicamente* possibili.

<sup>11</sup>La formulazione qui usata è leggermente diversa da quella di Kim; le modifiche apportate, puramente formali, possono rendere più facilmente confrontabili i quattro tipi di covarianza (l’ultimo dei quali dovuto a Terence Horgan, a cui però qualcuno attribuisce anche il terzo tipo).

E' chiaro che queste possibilità non soddisfano in alcun modo l'ipotesi fisicalistica della dipendenza mentale dagli eventi cerebrali.

Per quanto riguarda la condizione di *non-riducibilità*, Kim affronta prima di tutto il problema, che abbiamo esaminato all'inizio del capitolo, del carattere ambiguo del concetto di riduzione. In particolare, relativamente all'accezione nomologica di «riduzione», Kim osserva che, nell'esplicitarla, tutto dipende dal tipo di teorie coinvolte nella riduzione. Per aggirare tale difficoltà, egli considera una condizione più esplicita, che tuttavia garantisce la riduzione tra teorie: la “condizione di forte connettibilità” [Kim 1990, 151]. Nel caso di teorie formulate mediante un linguaggio del primo ordine, essa può essere enunciata come segue: ogni predicato P della teoria T *da ridurre* è connesso ad un predicato coestensivo Q della teoria *a cui ridurre* T, secondo una legge bicondizionale del tipo “ $\forall x, Px \leftrightarrow Qx$ ” (e analogamente per tutti i predicati relazionali). Se la sopravvenienza tra mente e cervello deve soddisfare la clausola di non-riducibilità, allora una relazione candidata ad essere tale sopravvenienza *non* deve comportare logicamente la forte connettibilità. Invece, sostiene Kim, ci sono buone ragioni per ritenere che la covarianza *forte* implichi la forte connettibilità<sup>12</sup>.

La situazione è così la seguente: la covarianza debole è troppo debole per costituire un nesso di dipendenza e quella forte è troppo forte per essere compatibile con la non-riducibilità. Se la sopravvenienza è una relazione di covarianza che unisce dipendenza e non-riducibilità, allora *né la covarianza debole né quella forte possono essere forme di sopravvenienza*.

Kim considera una terza possibilità (si ricordi che M è l'insieme delle proprietà sopravvenienti su F) [Kim 1990, 154]:

3. *Covarianza globale*: per ogni coppia di mondi (logicamente) possibili  $w$  e  $w'$ , se  $w$  e  $w'$  sono F-indiscernibili, allora sono M-indiscernibili.

---

<sup>12</sup>L'argomentazione di Kim a questo proposito è la seguente. Dalla covarianza forte segue che: se M è l'insieme delle proprietà sopravvenienti su F e se la proprietà P di M è istanziata da un oggetto  $x$ , allora c'è una proprietà Q di F tale che  $x$  ha Q e vale la legge “necessariamente se qualcosa ha Q, allora ha P”. Quindi, per ogni P di M, esiste un sottoinsieme di F  $\{Q_1, Q_2, \dots\}$ , tale che ciascuna  $Q_i$  appartenente a tale sottoinsieme è necessariamente sufficiente per l'istanziamento di P. Kim dimostra che P e la disgiunzione  $(Q_1 \vee Q_2 \vee \dots)$  sono coestensive e legate da una legge bicondizionale come quella richiesta dalla condizione di forte connettibilità [Kim 1990, 151-2].

A prima vista, questa relazione, applicata al rapporto mente-corpo, sembrerebbe catturare con precisione la nostra idea intuitiva di dipendenza senza riduzione. Ma Kim ha pronta la “doccia fredda”: la covarianza globale è, guardando meglio, debolissima. Essa, infatti, è compatibile con il seguente mondo possibile: un mondo completamente identico al nostro, in tutti i dettagli *fisici* possibili immaginabili, tranne che per un unico atomo di idrogeno isolato, che si trova in una posizione leggermente diversa da quella occupata da un identico atomo nel nostro mondo; ebbene, le facoltà mentali degli esseri di questo mondo possibile potrebbero essere, *compatibilmente con la covarianza globale*, diversissime da quelle degli esseri del nostro mondo<sup>13</sup> [Kim 1990, 155; 1989, 277].

A questo punto Kim conclude schierandosi per la covarianza forte, contro quella debole e quella globale, anche se accettare tale relazione come formalizzazione logica della sopravvenienza della mente sul cervello significa *rinunciare alla non-riducibilità*. In breve, la soluzione che Kim propone al problema mente-corpo è un fisicalismo *riduzionista* associato alla forma di riduzione che abbiamo chiamato “condizione di forte connettibilità”. Secondo tale prospettiva, l’unico modo di *spiegare perché la mente sopravviene sul cervello* è appellarsi alla presenza di relazioni di sopravvenienza *specifiche*, cioè appropriate leggi di correlazione tra *specifiche* proprietà mentali e proprietà cerebrali. Secondo Kim, se queste specifiche correlazioni sono esse stesse spiegabili, tanto meglio; *ma anche in caso contrario, invocarle dovrebbe costituire il primo passo necessario* [Kim 1990, 159].

Contro quest’ultima considerazione si è schierato con decisione Terence Horgan, il quale, in un articolo assai convincente [Horgan 1993], ha messo in guardia i filosofi della mente dall’utilizzazione di categorie esplicative a loro volta *non spiegabili*. Mentre Kim, tutto sommato, sembra ben disposto verso forme “forti” di sopravvenienza eventualmente inspiegabili, Horgan rifiuta l’idea che le relazioni di sopravvenienza possano essere *fondamentali*, cioè *sui generis* e inspiegabili.

---

<sup>13</sup>Questo perché, dal punto di vista logico, la condizione di covarianza globale (“se i due mondi sono fisicamente-indiscernibili, allora sono mentalmente-indiscernibili”) *non viene falsificata* dalla falsità sia dell’antecedente (“i due mondi sono fisicamente-indiscernibili”), sia del conseguente (“i due mondi sono mentalmente-indiscernibili”).



Ciò che i fisicalisti dovrebbero elaborare per formalizzare logicamente il rapporto mente-corpo è una sopravvenienza ontologica completamente spiegabile e spiegabile in un modo a sua volta materialisticamente spiegabile. Poiché finora nessuno è riuscito ad individuare una simile forma di sopravvenienza, Horgan ha battezzato tale elusiva relazione «*superdupervenience*» (“sopra-ingannevolvenienza”).

Horgan ha anche proposto una nuova direzione di indagine, la covarianza regionale [Horgan 1993, 571]<sup>14</sup>:

4. *Covarianza regionale*: per ogni coppia P e P' di regioni spazio-temporali di mondi (logicamente) possibili, se P e P' sono Q(F)-indiscernibili, allora sono Q(M)-indiscernibili<sup>15</sup>.

In questa sede non è possibile esaminare la (remota) possibilità che questa forma di covarianza sia proprio la *superdupervenience* che vorrebbe Horgan; si osservi solo, per inciso, che: (a) essa è molto più forte della covarianza debole; (b) contiene la covarianza globale come caso limite (se le regioni spazio-temporali sono mondi possibili interi).

A questo punto è necessario tirare le fila del discorso. Abbiamo visto che «fisicalismo» e «riduzionismo» sono termini che vanno precisati con grande attenzione, caso per caso. In certe accezioni, essi coincidono senza alcun dubbio. In altre accezioni, i filosofi hanno considerato la possibilità di un fisicalismo non-riduzionista. Il monismo anomalo di Davidson è, ad esempio, il tentativo di conciliare il fisicalismo dell'identità delle occorrenze al rifiuto del riduzionismo nomologico. Molti autori hanno tuttavia criticato il monismo anomalo, considerato da Kim addirittura indistinguibile dall'eliminativismo. Sempre Kim ha poi combattuto tutti i tentativi di utilizzare il concetto di sopravvenienza per sfuggire al riduzionismo *nomologico* adottando al contempo una prospettiva fisicalistica (nell'accezione di *dipendenza* dai fatti fisici).

---

<sup>14</sup>Horgan parla di sopravvenienza, non di covarianza. Per non generare confusione, è però preferibile inserire la sua proposta nel contesto del discorso di Kim [1990].

<sup>15</sup>Q(F) è il sottoinsieme delle proprietà *qualitative intrinseche* di F e, ovviamente, Q(M) è il sottoinsieme delle proprietà *qualitative intrinseche* di M (la notazione è di chi scrive). Le proprietà *intrinseche* di una regione spazio-temporale P sono quelle che non dipendono, nel più ampio senso della parola, da ciò che avviene fuori da P; le proprietà *qualitative* di P sono quelle che non dipendono, nel più ampio senso della parola, dall'esistenza di individui specifici entro P [Horgan 1993, 570-1].

Gli argomenti di Kim contro il fisicalismo non-riduzionista sono efficaci ma probabilmente non decisivi. Tuttavia, *se anche egli avesse completamente ragione a proposito del riduzionismo nomologico*, il discorso sarebbe ancora aperto relativamente ad altre interpretazioni del riduzionismo. In questa direzione si sono mossi ad esempio William Bechtel e Robert C. Richardson, che hanno: 1) criticato la focalizzazione del dibattito sulla riduzione nomologica; 2) proposto un'interpretazione del riduzionismo come posizione all'interno di una gamma di modalità di indagine.

1) Secondo Bechtel e Richardson, il modello nageliano di riduzione tra teorie è un mero esercizio filosofico [Bechtel - Richardson 1992, 259]. Tale modello, infatti, riguarda il rapporto tra teorie completamente sviluppate ed è quindi *a posteriori* della ricerca scientifica. Ciò non significa che esso *non potrebbe avere* un'efficacia euristica, ma che esso - di fatto - *non l'ha avuta* nel lavoro degli scienziati sperimentali. Comunque, il problema più grave che si ha con il modello di riduzione tra teorie è un altro.

Il vero problema è che in molti contesti scientifici, specialmente in biologia e in psicologia, la conoscenza non risiede in singole equazioni o in leggi condizionali (del tipo: "se x, allora y"), come quelle previste dal modello della riduzione nomologica (e *a fortiori* della "condizione di forte connettibilità" di Kim). In altre parole, la rappresentazione della conoscenza scientifica *reale* è solo in piccolissima parte logico-formale [Bechtel - Richardson 1992, 260]. Non solo: la ricerca sperimentale comporta la costruzione di teorie scientifiche che *non sono insiemi di leggi o equazioni* (soprattutto in biologia e in psicologia, dove i modelli sono spesso *qualitativi*). Infine, la forma di ragionamento coinvolta nella pratica dei ricercatori non è semplicemente una forma di logica induttiva. Per esempio, la soluzione ai due problemi "mereologici" - (a) identificare le parti di un sistema e (b) capire cosa fanno e come sono collegate - comprende varie manipolazioni del sistema ed è spesso non-verbalizzata (va dalle tecniche di visualizzazione, come la microscopia, alle procedure che modificano il sistema, come la centrifugazione) [Bechtel - Richardson 1992, 262].

*Tutte queste considerazioni suggeriscono di non condurre la discussione sul fisicalismo non-riduzionista semplicemente in base alla dicotomia “riduzione nomologica sì / riduzione nomologica no”.*

2) Il riduzionismo, sostengono i due autori, non solo non si esaurisce affatto con il riduzionismo tra teorie, ma in realtà è una *questione di grado* e non una scelta del tipo “tutto o niente”. Secondo Bechtel e Richardson, che *per riduzione intendono genericamente la ricerca della spiegazione del comportamento di un sistema nelle sue parti*, ogni approccio riduzionistico deve fare due assunti principali: la *modularità* e la *strutturazione gerarchica* del sistema studiato [Bechtel - Richardson 1992, 264]. La *modularità* di un sistema (da non confondere assolutamente con la modularità delle facoltà mentali ipotizzata da Fodor [1983]) corrisponde a quanto il sistema assomiglia ad un semplice aggregato<sup>16</sup>. La *strutturazione gerarchica* corrisponde invece alla seguente organizzazione sistemica: i costituenti fondamentali formano moduli (cioè sottostrutture stabili) interagenti, che a loro volta formano moduli più grandi, e così via; ad ogni livello, le interazioni intramodulari sono qualitativamente diverse da quelle intermodulari (e in genere le prime sono più forti delle seconde).

In teoria, inoltre, ogni sistema può essere guardato da due prospettive: quella della sua organizzazione *in natura* e quella della sua organizzazione *secondo il modello scientifico disponibile*. Purtroppo non c'è modo di conoscere scientificamente un sistema finché un suo modello non viene sviluppato, così in realtà l'unica prospettiva che si può assumere è la seconda [Bechtel - Richardson 1992, 266-7].

Dalle considerazioni precedenti segue che i (modelli scientifici dei) sistemi possono essere pensati lungo una *scala di modularità decrescente*, cioè una scala di importanza delle interazioni tra le parti *crescente*, dai semplici aggregati ai (modelli dei) sistemi le cui proprietà dipendono solo dalle interazioni tra i moduli e non dalle singole componenti. A questa scala, poi, si deve affiancare una *seconda scala*, relativa ai vari modi di studiare i sistemi, che va dal riduzionismo radicale all'olismo [Bechtel - Richardson 1992, 265-6]. E' ovvio che l'atteggiamento migliore da assumere dipende dalla collocazione, entro la prima scala, del sistema da studiare: se un sistema

---

<sup>16</sup>Le caratteristiche di un aggregato sono: l'interscambiabilità delle parti, la somiglianza qualitativa rispetto ai cambiamenti nel numero delle parti, la stabilità rispetto alla riaggregazione delle parti, interazioni minime tra le parti [Bechtel - Richardson 1992, 263].

è altamente modulare, l'atteggiamento più adatto è riduzionistico [Bechtel - Richardson 1992, 263-4]; ma se le proprietà sistemiche dipendono solo dalle interazioni tra le parti e non dalle parti stesse, allora occorre assumere un atteggiamento interessato meno alle parti e più al sistema nel suo insieme (olismo) [Bechtel - Richardson 1992, 266]. Per spiegare un fenomeno, dunque, non occorre tanto individuare le forze e i costituenti del livello "più basso", quanto piuttosto *identificare il giusto livello di analisi* [cfr. Bechtel - Richardson 1992, 265].

L'interpretazione del riduzionismo dovuta a Bechtel e Richardson permette di semplificare certi problemi filosofici di lunga data. L'antica controversia tra meccanicismo e vitalismo, per esempio, diventa meno insidiosa se considerata sullo sfondo della seconda scala. Inoltre, con la loro interpretazione, i due autori spostano il problema del riduzionismo dai *sistemi in sé* ai *livelli di analisi*, cioè dall'oggetto studiato al soggetto che lo studia. Nonostante la questione sia assai affascinante, in questa sede non possiamo soffermarci su di essa. Quello che qui ci interessa specialmente è sottolineare che lungo la scala degli atteggiamenti possibili verso un sistema, in posizione intermedia, tra riduzionismo e olismo, nonché tra meccanicismo e vitalismo, Bechtel e Richardson collocano l'*emergentismo* [Bechtel - Richardson 1992, 274].

Veniamo finalmente ad esso.