UNIVERSITÀ DEGLI STUDI DI SASSARI

DIPARTIMENTO DI SCIENZE UMANISTICHE E SOCIALI

SCUOLA DI DOTTORATO IN SCIENZE DEI SISTEMI CULTURALI

TEORIE E STORIA DELLE LINGUE E DEI LINGUAGGI

CICLO XXV

LA SOPRAVVENIENZA DEL MENTALE SUL FISICO: RIFLESSIONI SUL PROBLEMA MENTE-CORPO E IL RUOLO DELLE PROPRIETÀ MENTALI

Tutor:

Prof. Massimo Dell'Utri

Tesi di Dottorato di:

Silvia Negroni

INDICE

Introduzione 3

CAPITOLO 1. Le proprietà mentali

1.1. Le proprietà mentali 12

1.2. Il dualismo cartesiano 14
1.2.1. Il dualismo oggi 15

1.3. Il riduzionismo 19
1.3.1. Teoria dell'identità di tipo, o tipo-tipo 20
1.3.1.1. Alcune critiche alla teoria dell'identità di tipo 21
1.3.2. La teoria delle identità delle occorrenze 23

1.4. Il monismo anomalo 24
1.4.1. Gli eventi mentali 26
1.4.2. Le leggi psicofisiche 28
1.4.3. La sopravvenienza 30
1.4.4 La critica di Jennifer Hornsby 32
1.4.4.1 Il problema della transitività 37
1.4.4.2 Le azioni 38

1.5. Il funzionalismo 42
1.5.1. Il funzionalismo di Hilary Putnam 43
1.5.2. Funzionalismo e riduzionismo 48
1.5.3. Distacco dal funzionalismo 49
1.5.4 Il funzionalismo liberalizzato 52
1.5.5. Un'evoluzione del funzionalismo: la TCRM 53
1.5.6. Critiche al funzionalismo 56

1.6. Il punto focale del dibattito: la sopravvenienza 62

CAPITOLO 2. Jaegwon Kim: la sopravvenienza e il ruolo delle proprietà mentali

2.1. Le conseguenze della visione davidsoniana: epifenomenismo e eliminativismo 65
2.1.1. L'esclusione causale 66
2.1.2. La chiusura causale 68
2.1.3. Determinazione e covarianza 71

2.2. La soluzione di Kim 74
2.2.1. Le proprietà mentali 75
2.2.2. La sopravvenienza 79
2.2.3. La riduzione funzionale 81
2.2.4. Tre tappe per l'identificazione delle proprietà mentali 82

2.3. Alcune critiche 87

CAPITOLO 3. Hilary Putnam: la sopravvenienza globale

3.1. La posizione di Putnam 91

3.2. La critica a Kim 92
3.2.1. La causazione mentale 94
3.2.2. L'ipotesi degli automi: una condizione intellegibile? 96
3.2.3. La relazione tra proprietà mentali e proprietà fisiche: indipendenza e identità 97
3.3. L'esternismo semantico 99
3.4. Ulteriori riflessioni sull'ipotesi (SENZ'ANIMA) 101
3.5. Come definire e individuare una "credenza" 103
3.6. La falsa idea del mentale quale teatro interno 105
3.7. La sopravvenienza 107
3.8. Il realismo diretto (o naturale) 109

CAPITOLO 4. Stephen L. White: la a-sopravvenienza

4.1. La critica a Putnam 112
4.1.1 Cosa non va con la sopravvenienza 116
4.1.2 La risposta di Putnam 120

Conclusioni 122

Bibliografia 127

Ringraziamenti 140

Introduzione

Questa tesi riguarda le proprietà mentali e il concetto di sopravvenienza del mentale sul fisico. In particolare, essa si concentra sul modo in cui le proprietà mentali ci possono aiutare a capire come la nostra mente ci mette in relazione con l'ambiente naturale e sociale nel quale viviamo.

La ricerca si inserisce nel campo della filosofia della mente, un campo oggi molto vivo grazie anche ai progressi compiuti da discipline quali neurologia e psicologia, con le quali la filosofia della mente è fortemente legata.

La filosofia della mente si occupa di vari aspetti strettamente legati fra loro, tra i quali il rapporto mente-cervello (o, più in generale, mente-corpo); il modo in cui la mente riesce a rappresentare la realtà esterna; la previsione del comportamento; la semantica dei predicati mentali; la natura degli stati mentali; la teoria dell'azione; il problema delle altre menti e dell'intenzionalità degli stati mentali. Inoltre, è il concetto stesso di mente a sollevare diversi problemi, primo fra tutti quello della sua esistenza, ovvero se ci sia o no qualcosa di diverso dalla materia fisica rappresentata dal cervello.

Uno dei problemi studiati dai filosofi della mente, e argomento di questa trattazione, è quello relativo alle proprietà mentali. Quando ci si interroga sulle proprietà mentali inevitabilmente ci si imbatte in annose questioni filosofiche: una di queste questioni è il problema mente-corpo. Concetto centrale al riguardo è quello di *stato mentale*: uno stato mentale può essere un desiderio, un dolore, un'esperienza soggettiva, ecc. La vita mentale dunque è fatta di stati mentali. Le domande sugli stati mentali alle quali le principali tendenze in filosofia della mente tentano di dare una risposta riguardano, appunto, le proprietà degli

stati mentali e della mente. Ma non solo: infatti ci si interroga su come un sistema fisico come il cervello possa dare luogo a stati mentali che, evidentemente, non hanno niente di fisico. Le diverse concezioni della mente umana che si sono sviluppate in filosofia (concezioni dualiste o materialiste, e particolari declinazioni all'interno delle stesse teorie materialiste) portano inevitabilmente, come vedremo, ad una diversa caratterizzazione del concetto di proprietà mentale.

Ci siamo proposti tuttavia di limitare il campo di indagine e prendere come punto di riferimento autori quali Jaegwon Kim e Hilary Putnam passando per Donald Davidson e finendo con Stephen White, i cui lavori al riguardo ci sono parsi assai significativi.

Si metterà in evidenza che una nozione importante utilizzata da questi autori, e in particolare da Davidson e Kim, nello spiegare cosa sia una proprietà mentale e quale sia il suo posto all'interno della mente umana, è quella di "sopravvenienza". In breve, quanti affermano che le proprietà mentali *sopravvengono* su quelle fisiche, ritengono che la modifica di proprietà fisiche porta necessariamente alla modifica di proprietà mentali, ma non viceversa.

Di questa nozione tuttavia, come emergerà dalla ricerca, se ne sono date diverse interpretazioni a seconda di come viene concepita la realtà e il mentale; perciò ci è parso opportuno, prima di addentrarci nella spiegazione di questo concetto, offrire una panoramica di come sono state caratterizzate le proprietà mentali e a quali teorie queste diverse descrizioni hanno dato vita. Il primo capitolo del lavoro è dedicato pertanto alle posizioni in filosofia che hanno avuto maggiore rilevanza per gli autori succitati riguardo al discorso sulle proprietà mentali.

Per semplificare, in genere si tende a distinguere tra posizioni dualiste da una parte e posizioni materialiste dall'altra. Le prime hanno le loro basi nel dualismo cartesiano seicentesco, il quale sosteneva che la mente, vista come sostanza pensante e cosciente (la *res cogitans*), costituisce un tipo di realtà a sé, completamente indipendente dal corpo, inteso come sostanza fisica (la *res extensa*), e la cui esistenza quindi non dipende da quest'ultimo.

Il capitolo mette in luce anche gli sviluppi più recenti di questa posizione che, sebbene con modifiche più o meno rilevanti rispetto alla versione sostenuta da Cartesio, è oggi una tesi presente nel panorama filosofico. Giusto per citare alcuni nomi, personaggi quali John Eccles, Wilder Penfield, Karl Popper, David Chalmers e Thomas Nagel, seppur in modo diverso, sono dei dualisti.

Le posizioni opposte a quella dualista si collocano all'interno del cosiddetto "materialismo": la tesi fondamentale del materialismo (o fisicalismo) è che l'intera realtà è costituita da tutte le entità che sono riconosciute e postulate dalla fisica di base, considerata come la scienza fondamentale. Un mondo concepito come composto solo dalla totalità di tutti gli elementi fisici, ovviamente, non lascia spazio ad alcun elemento psicologico. Il problema è però che la natura degli stati mentali pare non essere materiale, perciò il problema risiede nel riuscire a coniugare la natura apparentemente non fisica del mentale con una visione materialistica del mondo. Questo tipo di visione è dunque già di per sé una posizione riduzionista, in quanto se nella realtà non c'è che materia fisica, è a questa che il mentale viene inevitabilmente ridotto.

Nel mostrare alcune proposte di soluzione a tale problema avanzate in filosofia della mente, abbiamo preso in considerazione solo quelle teorie che ci sono parse significative per la presente ricerca, ovvero le posizioni riduzioniste e il funzionalismo, lasciando fuori dal discorso tesi quali l'eliminativismo (per il quale la mente non esiste) e il comportamentismo (in base al quale la mente è un insieme di disposizioni al comportamento), tesi che hanno comunque svolto un ruolo importante all'interno del materialismo e della filosofia contemporanea della mente.

Per quanto riguarda il riduzionismo, verranno analizzate le due tesi dell'identità, ovvero la teoria dell'*identità di tipo* (*type*) e la teoria dell'*identità delle occorrenze* (*tokens*). I principali esponenti della prima corrente di pensiero, ovvero la teoria dell'identità di tipo, sono Herbert Feigl, Ullin Place e Jack Smart, i quali sostengono che per ogni stato mentale esiste un preciso corrispettivo stato fisico. L'identità alla quale si fa riferimento è dunque quella tra tutti gli stati mentali di un certo tipo (per esempio avere mal di denti) e tutti gli stati cerebrali di un certo altro tipo (per esempio avere le fibre C stimolate), ossia tra una certa proprietà mentale e una certa proprietà cerebrale.

Una versione più moderata della teoria dell'identità di tipo è la teoria dell'identità delle occorrenze, in base alla quale ogni singola occorrenza di proprietà mentale è identica a una corrispondente occorrenza di proprietà fisica; questa teoria pertanto ammette che, per esempio, due soggetti diversi possono trovarsi nello stesso stato mentale (per esempio credere che fuori splende il sole) e tuttavia trovarsi in differenti stati cerebrali. Ciò significa che allo stesso tipo di stato mentale (fuori splende il sole) non corrisponde sempre un particolare tipo di stato neuronale.

Questo tipo di teoria, vedremo, è alla base, del monismo anomalo di Donald Davidson: "monismo" poiché secondo Davidson gli eventi mentali sono eventi fisici; "anomalo" in quanto gli eventi, quando sono descritti in termini psicologici, non ricadono sotto leggi fisiche rigorose. Dunque in base al monismo anomalo un unico evento può essere descritto per mezzo del vocabolario delle scienze naturali o del vocabolario mentalistico, tutti gli eventi possono essere descritti nei termini della fisica e tutte le relazioni di causa ed effetto esemplificano leggi fisiche. Tuttavia quando gli eventi sono descritti in termini mentalistici non ricadono sotto leggi fisiche.

Il tipo di relazione che si determina tra mentale e fisico, secondo Davidson, è ciò che va sotto il nome di "sopravvenienza": con questo termine egli si riferisce al fatto che le caratteristiche mentali sono in qualche modo dipendenti dalle (o appunto sopravvenienti sulle) caratteristiche fisiche, ossia 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»[1]. Tuttavia, sostenendo Davidson una versione "debole" dell'argomento della sopravvenienza, egli sottolinea come questo tipo di relazione non vale, dal punto di vista ontologico, in tutti i mondi possibili.

La ricerca si propone di mettere in luce le problematiche che scaturiscono dalla caratterizzazione della causalità e della sopravvenienza nel monismo anomalo, ed è per questo che sono state prese in esame le critiche di Jennifer Hornsby e di Kim. Entrambi, seppur come vedremo utilizzando argomentazioni diverse, mettono l'accento sul problema della causalità nel monismo anomalo di Davidson. Infatti, la causalità così come intesa da quest'ultimo pare lasciare oscuro il tipo di relazione che sussiste tra la causa fisica e la causa mentale, dando un ruolo prioritario alla causa fisica senza lasciare spazio alla causa mentale.

L'ultima teoria che è stata analizzata nel primo capitolo è il funzionalismo che grande peso ha avuto nel pensiero di Putnam e che, anche se con qualche modifica, continua ad averlo; di orientamento funzionalista vedremo essere anche il pensiero di Kim. Il funzionalismo è una teoria che Putnam elabora negli anni '60 del secolo scorso e che costituisce una reazione al diffondersi di due prospettive entro la filosofia della mente: il comportamentismo logico (o filosofico) e il riduzionismo. La tesi centrale del funzionalismo, come vedremo, consiste nel fatto che gli stati mentali degli esseri umani possono essere individuati in modo totalmente indipendente dalla loro realizzazione fisica; questo porta a

[1] Davidson 1980, tr. it. p. 294.

identificarli in tutto e per tutto con quelli di una macchina calcolatrice, nella fattispecie la macchina di Turing, in quanto ciò che ha rilevanza non è la composizione fisica, bensì la forma. Infatti come l'identità di uno stato della macchina dipende dalle relazioni che esso ha con gli input, gli output, e gli altri stati del programma, analogamente si possono identificare gli stati mentali degli esseri umani in base al ruolo funzionale, il quale coincide con il ruolo causale dello stato, ossia con le relazioni che esso ha con input, output e altri stati mentali, indipendentemente dalla materia con i quali sono realizzati. Ciò che conta quindi è l'organizzazione o la struttura logica del sistema in questione, prescindendo appunto dalla sua realizzazione fisica.

Vedremo tuttavia come Putnam stesso, dopo aver fondato tale corrente di pensiero, se ne allontani, mantenendo ferma tuttavia l'idea che ciò che conta per le proprietà mentali è la giusta capacità funzionale, ma queste funzioni non sono esclusivamente interne e non sono descrivibili nel vocabolario della scienza: «although I am no longer a *computer-program* functionalist, the idea that *our mental states are best conceived of as ways of functioning and exercises of those ways* still seems right to me, although not in the "internalist" (and reductionist) sense that went with the model of those states as "the brain's software"»[2].

Il secondo capitolo si apre con la critica di Kim nei confronti di Davidson. Kim fondamentalmente accusa il monismo anomalo e la correlata nozione di sopravvenienza di portare a modalità scorrette di riduzione fra proprietà mentali e fisiche e quindi ad annullare qualsiasi potere causale delle proprietà mentali (*epifenomenismo*) o a eliminare completamente queste proprietà stesse (*eliminativismo*).

Kim basa le sue argomentazioni sulla presunta violazione che il monismo anomalo farebbe di due principi che egli definisce invece fondamentali per poter sostenere una qualsiasi posizione materialista, ovvero il principio di *esclusione causale* e il principio della *chiusura causale*.

Ciò che emergerà è che le accuse di Kim riguardano il ruolo causale del mentale nel monismo anomalo, ruolo che per Kim diventa nullo una volta che si dimostra come per il monismo anomalo due eventi per essere relazionati causalmente devono esemplificare una legge fisica, quindi gli eventi hanno un ruolo causale quando sono descritti fisicamente; sembra dunque che sia grazie alle proprietà fisiche, e non a quelle mentali, che gli eventi esemplificano le leggi della fisica. Pare così che dietro al presunto ruolo causale giocato dalle

[2] Putnam 2012, cap. 36 *How to Be a Sophisticated "Naïve Realist"*, p. 625.

proprietà mentali ci sia sempre una proprietà fisica che esemplifica tale ruolo: è solo in virtù delle loro proprietà fisiche che gli eventi possono avere efficacia causale, gli eventi mentali, in quanto tali, non causano alcunché.

Dopo aver illustrato i punti deboli che Kim identifica nel monismo anomalo, presenteremo la soluzione avanzata da quest'ultimo nel caratterizzare la relazione tra mentale e fisico. Il modello di sopravvenienza proposto da Kim è quello della causazione sopravveniente, o sopravvenienza forte, il quale trova le sue fondamenta nella *riduzione funzionale*, in base alla quale le proprietà sopravvenienti sono causalmente efficaci se e solo se sono riducibili funzionalmente a proprietà precedentemente accettate come causalmente efficaci. In base a questa caratterizzazione della nozione di sopravvenienza una proprietà mentale sopravviene alla proprietà fisica corrispondente perché quella proprietà fisica la realizza.

La riduzione funzionale è dunque la via che Kim propone per identificare una proprietà mentale con i suoi realizzatori fisici e che egli suddivide in tre tappe fondamentali:

1. l'interpretazione delle proprietà che devono essere ridotte in base al ruolo causale che hanno;
2. la ricerca dei realizzatori fisici delle proprietà funzionali;
3. l'identificazione dei ruoli con i realizzatori.

Questa visione, e le conseguenze alle quali porterebbe, è stato oggetto di interessanti critiche da parte di Putnam, critiche che verranno analizzate nel terzo capitolo. In particolare Putnam analizza la lettura che Kim dà di Davidson, mettendone in luce i punti critici, dovuti a quella che Putnam ritiene essere una erronea concezione del mentale. Egli sottolinea come l'errore nel caratterizzare le proprietà mentali come fa Kim parta dal fatto di vederle come interne al nostro corpo, come qualcosa che sta dentro le nostre teste e che per questo ci sia bisogno di capire come sono relazionate con il nostro cervello.

Putnam invece ritiene che si debba considerare la mente come un sistema di abilità funzionali, come un complesso di capacità conoscitive (materiali e non), dove non ci sono proprietà già date, e dove la realtà è da rinegoziare man mano che il nostro linguaggio e la nostra vita si sviluppano. Il pensiero dell'autore è fortemente influenzato dall'esternismo semantico, ovvero dall'idea che la connessione tra il mondo e i nostri pensieri sia necessaria perché i fattori esterni al soggetto contribuiscono in modo essenziale alla determinazione dei contenuti mentali.

Se non teniamo in considerazione questa connessione tra il mondo esterno e il nostro pensiero rischiamo di dare vita a posizioni filosofiche del tutto inaccettabili, sottolinea Putnam, come per esempio quelle legate all'idea cartesiana del mentale quale "teatro interno", ovvero del mentale come qualcosa che sta dentro alle nostre teste senza nessuna connessione con ciò che sta al di fuori di esse. La posizione difesa da Kim è per Putnam un chiaro esempio dell'erronea concezione delle proprietà mentali intese in questo senso.

A tal proposito si evidenzierà che il motivo per il quale Putnam rifiuta con tanta forza questa caratterizzazione delle proprietà mentali è che egli è fermamente convinto che sia proprio l'errore nel concepire il mentale come teatro interno, come interfaccia cognitiva tra la mente e il mondo, a dare vita a numerosi problemi all'interno della filosofia della mente (e non solo). L'assenza di un'interfaccia cognitiva tra noi e la realtà esterna implica che la nostra relazione con il mondo sia diretta, e questo porta al fatto che non ci sia uno stato mentale interno.

Putnam quindi respinge l'idea che la tesi della sopravvenienza, cioè della correlazione psicofisica, abbia un senso, poiché presuppone un'immagine sbagliata, ovvero quella secondo cui le nostre caratteristiche psicologiche sono stati interni correlati con ciò che accade dentro ai nostri corpi.

Tuttavia sottolineeremo come Putnam non rifiuta completamente la nozione di sopravvenienza, accettando un'idea di "sopravvenienza globale" che non si riferisce alla correlazione tra stati mentali e stati fisici, bensì all'idea che tutte le nostre capacità di esseri umani sopravvengono sugli stati dell'universo fisico, e non del soggetto.

Una delle critiche che ci è parsa più incisiva nei confronti di Putnam, e per questo motivo è stata oggetto di analisi, è stata quella di Stephen White. White sottolinea come forse il pensiero putnamiano non sia immune dall'accusa di porre un'interfaccia tra mente e mondo, come egli invece pretende sia. Abbiamo detto che Putnam critica fortemente l'idea che esista un'interfaccia cognitiva tra i nostri pensieri e il mondo esterno, tuttavia si vedrà come la critica di White faccia vacillare questa convinzione, almeno nei termini intesi da Putnam. Inoltre la critica di White si estende anche alla nozione di sopravvenienza globale sostenuta da Putnam, come conseguenza dell'argomento trascendentale difeso da White, in base al quale, riassumendo, i pensieri che riguardano gli oggetti sono necessari per avere un linguaggio significante e il fatto che riguardano gli oggetti consente che la relazione dei nostri pensieri con il mondo sia diretta.

Mostreremo come tale argomento porti a considerazioni che ci sono parse di grande rilevanza sia relativamente alla nozione di sopravvenienza sia, di conseguenza, relativamente alla questione mente-corpo. White, partendo dal presupposto che mentale e fisico sono incommensurabili, difende l'idea della *a-sopravvenienza*, ovvero l'idea che non c'è un modo di vedere qualcosa di comune tra mentale e fisico, per cui il problema stesso di come essi possano interagire perde di qualsiasi significato. Metteremo dunque in evidenza come queste riflessioni di White sembrano fornire una morale generale, convincente e plausibile, riguardo al dibattito sulla sopravvenienza che abbiamo illustrato nella ricerca.

CAPITOLO 1

Le proprietà mentali

1.1 Le proprietà mentali

Prima di addentrarci nel discorso su come sono state caratterizzate le proprietà mentali, ci pare opportuno dare alcune brevi delucidazioni sul concetto stesso di proprietà in generale e di proprietà mentale in particolare, ricordando che la visione che i filosofi di volta in volta hanno sulla natura delle proprietà stesse è influenzata dai problemi che le proprietà si pensa possano risolvere.

Una proprietà, in genere, viene definita come attributo di un oggetto: le proprietà infatti sono quelle caratteristiche che possono essere predicate di un oggetto, quelle che gli si possono appunto attribuire. Per esempio, se diciamo che quella cosa là è una mela ed è rossa, stiamo attribuendo la proprietà *rosso* e la proprietà *mela* a quell'oggetto.

Una questione importante riguardo le proprietà, oltre al fatto se esistano o meno – ma qui per semplicità stiamo supponendo che esistano –, è se esse siano universali o particolari (o *tropi*). Dire che una proprietà è universale significa sostenere che può essere istanziata da diverse cose; per tornare all'esempio di prima, è possibile per mele differenti esemplificare lo stesso colore, poniamo rosso, che è un singolo universale. Le proprietà universali «are entities that can be (and often are) literally shared by objects. It is possible for wholly distinct, non-overlapping objects to possess one and the same property at the same time»[3]. Le proprietà universali sono dunque proprietà astratte che possono essere condivise in diversi luoghi e in tempi diversi, caratterizzano oggetti diversi ma al contempo offrono una descrizione che li accomuna per una qualche loro somiglianza.

[3] Heil e Robb 2003, p. 177.

La visione opposta è quella che vede le proprietà come dei particolari astratti che possono caratterizzare un solo oggetto per volta e che possono essere interamente presenti in oggetti o individui che presentano tale proprietà, ma logicamente incapaci di essere al tempo stesso interamente presenti altrove[4]. Sempre per riferirci allo stesso esempio, non importa quanto simile possa essere il colore di due mele diverse, i loro colori sono proprietà distinte: il rosso della prima mela e il rosso della seconda mela.

La differenza tra universali e particolari è fondamentale in filosofia della mente, soprattutto se relazionata al discorso della causalità del mentale e la sostenibilità del riduzionismo: «our theoretical decisions at this foundational level may have far-reaching consequences for issues in this discipline, including the contrast between reductive and nonreductive physicalism, multiple realizability and mental causation»[5].

Collegata a questo discorso è la caratteristica di alcune proprietà di essere riconosciute per i poteri causali che conferiscono quando vengono istanziate. Se è ampiamente riconosciuto che le proprietà fisiche hanno poteri causali, lo stesso non vale per le proprietà mentali.

Le attuali discussioni in filosofia della mente a volte si riferiscono anche a livelli, o strati, di proprietà; si suppone quindi che esistano diversi livelli tra le proprietà. Da tale punto di vista, le proprietà microfisiche, ossia quelle che si riferiscono a oggettive caratteristiche del mondo naturale, sono viste come proprietà di basso livello, mentre le proprietà mentali, ovvero quelle proprietà che dipendono dalla mente, sono invece viste come proprietà di alto livello.

Quando si discute di proprietà mentali, tra l'altro, spesso si parla di proprietà qualitative (o *qualia*): queste includono le proprietà delle sensazioni corporee, delle emozioni, dell'esperienza percettiva, insomma tutte quelle proprietà di cui si ha consapevolezza solo quando se ne fa esperienza in prima persona. Queste spesso sono le proprietà che creano più problemi ai filosofi della mente, proprio perché legate all'esperienza in prima persona.

Un altro aspetto che viene discusso quando affrontiamo il discorso mentale sono le proprietà intenzionali, come per esempio avere la credenza che oggi sia una bella giornata, ovvero quelle proprietà che sono dotate di un contenuto che coinvolge l'ambiente. Spesso le

[4] Cfr. Robb 1997, p. 186.
[5] Gozzano e Orilia 2008, p. 7; cfr. anche Robb 1997.

proprietà qualitative sono viste come interne al nostro corpo, viceversa le proprietà intenzionali sono viste come relazionate con il mondo esterno.

La questione principale che viene discussa è, quindi, se le proprietà mentali sono qualcosa di diverso dalle proprietà fisiche o se invece sono riducibili o almeno in qualche modo riconducibili ad esse. E se sono qualcosa di diverso, allora si discute su quale ruolo hanno nella nostra vita. Vedremo di seguito alcune risposte che la filosofia ha tentato di dare a questo problema.

1.2. Il dualismo cartesiano

Quando si affronta il problema delle proprietà mentali si devono fare i conti con le diverse teorie che si sono occupate di esse e indirettamente della questione mente-corpo.

Tale problema, molto antico nella filosofia, generalmente si fa risalire a René Descartes e al cosiddetto "dualismo cartesiano". In pieno Seicento, infatti, Descartes propose una tesi secondo la quale la mente costituisce un tipo di realtà a sé, completamente indipendente dal corpo e la cui esistenza quindi non dipende da quest'ultimo.

Descartes partiva dal presupposto che la sostanza pensante e cosciente (la *res cogitans*) fosse contrapposta e irriducibile alla sostanza fisica (la *res extensa*). Egli riteneva infatti che il cosiddetto "argomento del *cogito*" (se dubito di qualcosa e quindi penso, allora necessariamente esisto, da qui *cogito ergo sum*) rivelasse indirettamente che solo il pensiero è la proprietà essenziale dell'io: a prescindere da qualunque altra cosa potesse essere, esso era innanzitutto "cosa pensante":

> io trovo che il pensiero è un attributo che mi appartiene: esso solo non può essere distaccato da me. […] Io non sono […] se non una cosa che pensa, e cioè uno spirito [una mente], un intelletto o una ragione […] Ma che cosa sono dunque io? Una cosa che pensa. E che cosa è una cosa che pensa? Una cosa che dubita, che concepisce, che afferma e che nega, che vuole, che non vuole, che immagina, anche e che sente[6].

Pensare era l'unica attività che l'uomo poteva svolgere senza bisogno del corpo, e secondo Descartes questo sarebbe provato dal fatto che posso immaginare di non avere un

[6] Descartes 1954, trad. it. pp. 80-81.

corpo e comunque continuare a esistere, tuttavia non sarebbe possibile il contrario, ovvero una volta che smetto di pensare, io smetto anche di esistere. E se per l'io solo il pensare, e non il corpo, è essenziale, allora, secondo Descartes, abbiamo dimostrato che la mia mente e il mio corpo sono due sostanze differenti. Mente e corpo non possono coincidere perché la mente non è divisibile, è «una realtà assolutamente una e intera»[7], mentre per il corpo non è così; infatti, se «una qualsiasi parte del corpo viene amputata, riconosco chiaramente che nulla, con questo, viene tolto alla mente»[8].

Tuttavia rimaneva il problema di come fosse possibile che due sostanze radicalmente diverse potessero essere unite, come spiegare le loro evidenti interazioni. Lo stesso Descartes venne accusato di rendere inintelligibile la causalità mentale[9]. Per questo Descartes avanzò la tesi che mente (o anima) e corpo non sono separati come nocchiero e nave, dove se la nave ha uno squarcio il nocchiero ne ha sì consapevolezza però non in modo immediato, ma sono intimamente legati e addirittura mischiati. Questo sarebbe evidente dalle sensazioni fondamentali quali la fame, la sete, il dolore: so in modo immediato e diretto, tramite il dolore appunto, se il mio corpo è ferito.

Esisterebbe quindi un punto dove mente e corpo interagiscono, dove *res cogitans* e *res extensa* entrerebbero in contatto e si scambierebbero informazioni attraverso gli "spiriti animali" (i nostri impulsi nervosi) che sono prodotti dalle percezioni esterne e che a loro volta producono il movimento del corpo: questo punto è la ghiandola pineale.

La tesi dunque sostenuta da Descartes è che mente e corpo sono entità separate che interagiscono causalmente, ovvero gli stati mentali causano stati fisici, sia nel nostro corpo che nel mondo esterno: voglio alzare un braccio e il braccio si alza (dunque la mente è causa di mutamenti corporei); gli stati fisici causano stati mentali: mi rompo una gamba e provo dolore (quindi il corpo è causa di fenomeni mentali). Per questo il dualismo cartesiano viene definito interazionistico, poiché mente e corpo sono sì distinti, ma la loro unione è comunque strettissima.

1.2.1 Il dualismo oggi

[7] Ivi, p. 263.
[8] *Ibid*.
[9] Crane 2001, trad. it. pp. 58-63. Crane ritiene che il problema non sia l'idea in sé di causare qualcosa mentalmente, ma che sia problematico il fatto di dover mettere assieme asserzioni non controverse sulla causalità con una tesi controversa sulla mente, ovvero il non avere collocazione spaziale.

Al giorno d'oggi, anche grazie ai progressi fatti dalle scienze, il dualismo così come inteso da Descartes è diventata una tesi difficile da sostenere, sebbene se ne trovano ancora tracce. Proprio a causa dei cambiamenti che ha subito nel corso del tempo, il termine "dualismo" viene oggi diviso in *ontologico* ed *epistemico*[10].

Con *dualismo ontologico* si intende una tesi molto vicina al dualismo originario di Descartes, secondo la quale la mente è sostanza separata dal mondo fisico, priva di dimensione temporale, indefinita, libera e inestesa. Esso si divide ulteriormente in:

- *dualismo delle sostanze*: corrispondente al dualismo cartesiano classico, il quale vede mente e cervello come due sostanze diverse;
- *dualismo delle proprietà*: la sostanza può essere la stessa, ma il pensare e la materialità sono aspetti completamente separati e irriducibili l'uno all'altro.

I maggiori esponenti contemporanei del dualismo ontologico sono filosofi come John Eccles e Wilder Penfield che si definiscono appunto dualisti[11]. Costoro difendono la peculiarità dell'aspetto fenomenico (o qualitativo) del mentale. Per loro gli stati mentali, come stati di esperienza, non si possono ridurre a meri stati fisici perché gli stati mentali sarebbero caratterizzati da alcune proprietà estranee al mondo fisico. Infatti, come afferma Eccles, «più cose scopriamo sul cervello, più siamo in grado di distinguere con chiarezza gli eventi cerebrali dai fenomeni mentali»[12]: per lui appunto «la neurologia clinica e le neuroscienze hanno chiarito in maniera sufficiente che la mente non ha accesso diretto al corpo. Tutte le interazioni col corpo sono mediate dal cervello»[13].

Un altro importante dualista dell'età moderna è Karl Popper[14] il quale, insieme al già citato Eccles, formula una teoria sulla natura della mente e sul suo rapporto con il cervello basandosi sul riconoscimento di tre mondi, che corrispondono a tre livelli ontologici della realtà: il Mondo 1 degli stati ed eventi fisici; il Mondo 2 degli stati ed eventi mentali; il Mondo 3 dei contenuti oggettivi di pensiero. Tra essi intercorrono relazioni di causa ed effetto dove i mondi superiori sono capaci di retroagire su quelli inferiori:

> col produrre una nuova opera, uno scultore può incoraggiare altri scultori a copiarla o a produrre sculture simili. La sua opera – non tanto attraverso i suoi aspetti

10 Cfr. Nannini 2002 e De Caro 1998.
11 Cfr. Paternoster 2002, p. 7.
12 Eccles e Robinson 1984, tr. it. p. 47.
13 Ivi, p. 44.
14 Popper e Eccles 1977.

> materiali quanto attraverso la nuova forma che ha creato [Mondo 3] – può influenzarli tramite le loro esperienze del Mondo 2 e, indirettamente, attraverso il nuovo oggetto del Mondo 1[15].

Una delle critiche che vengono rivolte a tale impostazione, e un po' a tutto il dualismo in generale, è quella di violare il primo principio della termodinamica, ovvero il principio di conservazione dell'energia (la quantità di energia contenuta all'interno di un sistema isolato è costante). Partendo dai presupposti che l'energia non si genera e non si distrugge, il primo principio della termodinamica afferma che, sebbene possa essere trasformata e convertita da una forma all'altra, la quantità totale di energia di un sistema isolato (come l'universo e il corpo umano) è una costante, ovvero il suo valore si mantiene immutato al passare del tempo. Per esempio, quando una palla da biliardo ne urta un'altra, è la palla che colpisce che cede a quella colpita l'energia cinetica necessaria per metterla in moto, la quantità di energia presente nell'universo prima e dopo il contatto (considerando anche quell'energia che si è trasformata in calore nell'urto stesso), rimane invariata. Normalmente l'energia necessaria per produrre un effetto fisico proviene dalla sua causa fisica (come nelle palle da biliardo). Dunque tornando al nostro discorso, se gli eventi mentali non hanno, per ipotesi, alcuna causa fisica (poiché se l'avessero sarebbero determinati da essa e non dalla mente) ma producono degli effetti che sono fisici e che quindi richiedono un consumo di una certa quantità di energia, allora l'energia che essi consumano non proviene dalla trasformazione di altra energia già presente nel cervello. Tale energia si aggiungerebbe a quella già presente nell'universo fisico prima del loro accadere, e questo però viola il primo principio della termodinamica[16].

Difensore del dualismo delle proprietà è David Chalmers[17], il quale ritiene che, sebbene sia il cervello a guidare il nostro comportamento, gli stati mentali non sono ontologicamente riducibili al cervello stesso. La mente non interferisce con il funzionamento del cervello, ma non è comunque riducibile a esso. Nel tentativo di dimostrare l'irriducibilità della coscienza rispetto ai processi nervosi del cervello Chalmers, rifacendosi ad autori quali Saul Kripke, Thomas Nagel e altri, elabora l'argomento della possibilità logica dell'esistenza degli *zombie*. Egli è del parere che è possibile, da un punto di vista strettamente logico,

15 Ivi, p. 55.
16 Cfr. Nannini 2002, p. 151.
17 Chalmers 1996.

ammettere l'esistenza degli zombie, ovvero di esseri in tutto simili agli uomini dal punto di vista fisico, ma privi di coscienza. Egli ritiene che sia logicamente possibile l'esistenza di un mondo fisicamente del tutto identico al nostro però privo di coscienza. Se è possibile immaginare mondi possibili fisicamente indistinguibili dal nostro, ma nei quali la coscienza è assente, allora questo significa che la coscienza, o gli stati mentali, non possono essere stati fisici. E questa, secondo Chalmers, sarebbe la prova della non riducibilità delle esperienze coscienti ai fenomeni nervosi che si svolgono nel cervello: la coscienza appartiene al mondo naturale, non è qualcosa di fisico e per spiegarla occorre trovare delle leggi psicofisiche le quali connettano le proprietà dei processi fisici con le proprietà dell'esperienza. Poiché l'esperienza dipende sì da processi fisici, ma questa dipendenza non può essere derivata dalle sole leggi fisiche, le leggi psicofisiche fungono da "ponte esplicativo" per spiegare come l'esperienza sorge da processi fisici ma non è a essi riducibile. Per Chalmers dunque la possibilità logica che esistano degli zombie è la prova dell'irriducibilità della coscienza a processi cerebrali. Gli zombie, infatti, sono fisicamente identici all'uomo, ma non sono dotati come l'uomo di una coscienza, la quale può trovare una spiegazione all'interno delle scienze naturali.

Questo tipo di dualismo viene definito "naturalistico", poiché la fisica non è più la scienza prima che descrive l'intero mondo reale, e la coscienza si inserisce tra gli elementi irriducibili della realtà, e sostiene che questa può essere spiegata all'interno delle scienze della natura.

Le critiche che vengono mosse a questa teoria di Chalmers vanno dal *panpsichismo*, secondo il quale anche la materia fisica ha un potenziale cosciente (da cui deriverebbe che anche un termostato sarebbe in grado di avere una qualche esperienza cosciente), all'*epifenomenismo*, il quale ammette la possibilità di esseri che, seppur privi di coscienza, possano comportarsi esattamente come degli esseri umani, a patto di negare alla coscienza qualsiasi rilevanza causale nel mondo dei fenomeni fisici, e di considerarla quindi un semplice epifenomeno.

Con *dualismo epistemologico* invece si intende quella concezione in base alla quale la sfera mentale viene "conosciuta" in modo diverso da come viene esperita la sfera della natura; Mario De Caro ne sottolinea tre aspetti fondamentali:

1. l'indipendenza costitutiva dei contenuti mentali, i quali possono essere causati da eventi ed oggetti esterni ma i cui contenuti non sono costituiti da una relazione con il

mondo esterno al soggetto, anzi sono considerati indipendenti da ogni condizione esterna alla mente;

2. il primato epistemico della conoscenza in prima persona, ovvero il fatto che il soggetto gode di un accesso privilegiato nella conoscenza dei propri stati mentali;
3. la rilevanza della problematica scettica, secondo la quale la realtà esterna al soggetto potrebbe essere per lui remota, inattingibile[18].

Un esponente di questo tipo di dualismo è Thomas Nagel il quale si sofferma soprattutto sul secondo punto. Egli ritiene che ogni fenomeno soggettivo è essenzialmente connesso con un singolo punto di vista. Nel suo celebre esperimento Nagel si domanda cosa si prova a essere un pipistrello, mettendo in luce come solo un pipistrello può rispondere a questa domanda, poiché, seppur studiamo il funzionamento del suo sonar, non si sa che cos'è l'esperienza di essere un pipistrello, che *effetto fa* essere un pipistrello. Molti processi mentali possono essere conosciuti adeguatamente e completamente solo "dall'interno" o "in prima persona", cioè solo se si è il soggetto in cui si svolgono quei processi. Nagel ritiene che «un organismo possiede stati mentali coscienti se e solo se si prova qualcosa a *essere* quell'organismo: se *l'organismo* prova qualcosa a essere quello che è»[19].

Per concludere il discorso sul dualismo va notato che una delle sue tesi più forti – ovvero il fatto che le nostre credenze sul mondo extramentale vengono giustificate a partire dalla conoscenza in prima persona, che l'intelletto umano è capace di conoscere in modo diretto solamente il contenuto delle proprie idee e soltanto attraverso esse può arrivare a conoscere in maniera indiretta gli oggetti esterni – ha influenzato notevolmente l'empirismo[20].

1.3. Il riduzionismo

In tempi più recenti, in seguito anche agli sviluppi compiuti da discipline quali neurologia e psicologia, vengono elaborate in filosofia diverse teorie di stampo materialistico. La tesi fondamentale del materialismo (o *fisicalismo*) è che l'intera realtà è costituita da massa ed energia, e da tutto ciò che comprende la fisica di base, considerata come la scienza fondamentale. Il problema è però che la natura degli stati mentali non è materiale: si tratta perciò di riuscire a coniugare la loro natura con una visione materialistica del mondo.

[18] De Caro 1998, pp. 127-35.
[19] Nagel 1974, tr. it. p. 380.
[20] Cfr. De Caro 1998, p. 135; Nannni 2002, p. 25.

In quel che segue si prenderanno in considerazione solo quelle teorie che ci paiono significative per la ricerca che si vuole effettuare. Analizzeremo perciò le posizioni riduzioniste e il funzionalismo, lasciando fuori dal discorso tesi quali l'eliminativismo (il quale sostiene che la mente non esiste) e il comportamentismo (il quale sostiene che la mente è un insieme di disposizioni al comportamento), le quali hanno comunque avuto un peso notevole all'interno del materialismo e della filosofia contemporanea della mente.

Per il riduzionismo, corrente che si impose durante gli anni '50 del secolo scorso, la mente è uguale al cervello e questo comporterebbe una riduzione degli stati mentali agli stati cerebrali. La mente perciò viene vista come un prodotto dell'attività cerebrale.

Tuttavia è importante non confondere l'idea di riduzione con quella di eliminazione, come ci fa notare T. Crane il quale, riprendendo le parole di Huw Price, sottolinea che «una riduzione identifica le entità di un dominio con una sottoclasse delle entità di un altro»[21].

Secondo il riduzionismo, e da qui il nome, è possibile tradurre, o meglio ridurre, la classificazione psicologica degli stati mentali nei termini delle proprietà fisico-biologiche del sistema nervoso. L'ambito del mentale può essere ridotto a quello del fisico, ovvero i concetti mentalistici che utilizziamo nel descrivere gli eventi mentali sono riducibili a concetti fisicalistici. Sarebbe perciò possibile una caratterizzazione degli stati mentali che passa attraverso una *riduzione interteorica*, basata su una legge-ponte che consente di ridurre una teoria di alto livello a una di basso livello. La riduzione interteorica consiste infatti nel passaggio da una teoria di alto livello T_1 (per esempio la psicologia) a una teoria di livello T_2 (per esempio la neurologia); questa riduzione è però possibile se e solo se tutti i predicati di T_1 possono essere tradotti nel vocabolario di T_2, una traduzione a sua volta possibile grazie a leggi-ponte, le quali sono enunciati che mettono in correlazione le proprietà di T_1 con quelle di T_2. In tal modo, le leggi-ponte fondamentalmente connettono la psicologia e la neuroscienza, cosa che inserirebbe la psicologia tra le scienze della natura. Quindi, in base a questo principio, si potrebbe ridescrivere uno stato postulato da una teoria psicologica di alto livello nei termini del livello fisico, ossia delle proprietà di livello neuronale. Abbiamo così un intreccio del piano ontologico, o metafisico, con quello epistemologico.

Il riduzionismo si differenzia in due teorie dell'identità: la teoria dell'*identità di tipo* (*type*) e la teoria dell'*identità delle occorrenze* (*tokens*)[22]. Vediamole nel dettaglio.

---

21 Crane 2001, p. 78.

22 La terminologia è ripresa da C.S. Peirce. Cfr. Paternoster 2002, p. 9.

1.3.1 Teoria dell'identità di tipo, o tipo-tipo

I principali esponenti di questa corrente di pensiero, come Herbert Feigl[23], Ullin Place[24] e Jack Smart[25], sostengono che per ogni stato mentale esiste un preciso corrispettivo neurologico; inoltre, poiché la mente non sarebbe che il prodotto dell'attività cerebrale, lo stato mentale sarebbe identico a tale corrispettivo. L'identità alla quale si riferisce la teoria è quella tra tutti gli stati mentali di un certo tipo (per esempio avere mal di denti) e tutti gli stati cerebrali di un certo altro tipo (per esempio avere le fibre C stimolate), ossia tra una certa proprietà mentale e una certa proprietà cerebrale. La teoria dell'identità dei tipi prevede quindi una riduzione dei concetti mentali ai loro equivalenti fisici, identificati con proprietà del cervello; le proprietà mentali sono identiche a certe proprietà fisiche: uno stato mentale (M) è la stessa identica cosa di uno stato fisico (F).

Da questo deriverebbe il nome teoria dell'identità di tipo (o tipo-tipo): ciò che è fondamentale è l'identificazione delle proprietà mentali con quelle fisiche.

Questo tipo di identità sarebbe del tutto analogo a quella che intercorre tra temperatura ed energia cinetica molecolare media: la temperatura, ovvero una proprietà fenomenica, altro non è che la risultante del moto delle molecole che formano il gas, ovvero una proprietà fisica. Quindi sono classificazioni terminologiche diverse dello stesso fenomeno, esattamente come dolore e fibre C.

Secondo Alfredo Paternoster, il pregio di questa teoria sarebbe quello di eliminare il dualismo, in quanto gli stati mentali possono causare gli stati fisici proprio perché sono stati fisici.

1.3.1.1 Alcune critiche alla teoria dell'identità di tipo

Diverse critiche sono state mosse a questa teoria, la quale ormai ha perso gran parte del suo prestigio.

[23] Feigl 1967.
[24] Place 1956.
[25] Smart 1959.

Una prima critica si fonda sul principio noto come "legge di Leibniz", in base al quale se due oggetti sono identici, allora condividono tutte le proprietà. Dunque basterebbe dimostrare l'esistenza anche di una sola proprietà posseduta dai tipi di eventi mentali, ma non dai corrispettivi tipi di eventi fisici, per confutare la teoria dell'identità. E questo appare assai probabile, soprattutto quando si prende in considerazione l'intenzionalità degli stati mentali (ovvero il fatto che essi si riferiscono a qualcosa, che sono dotati di un contenuto[26]), aspetto che invece è assente negli stati fisici.

Tuttavia una delle critiche che viene mossa a questa teoria, e che ha avuto maggior peso, viene dai funzionalisti, e in particolare da Hilary Putnam[27]. Egli afferma che il fatto che una proprietà mentale sia identica a una proprietà cerebrale sarebbe un'assunzione molto vincolante. Se si parte, per esempio, dal presupposto che si attribuiscono stati mentali anche agli animali, ci si troverebbe nell'improbabile situazione che, nonostante uomini e animali abbiano cervelli diversi, quando condividono uno stato mentale si trovano nello stesso identico stato cerebrale:

> Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say 'hungry'), but whose physical-chemical 'correlate' is different in the two cases, the brain-state theory has collapsed. It seems to me overwhelmingly probable that we can do this[28].

Secondo questa obiezione quindi, che prende il nome di "realizzabilità multipla", una stessa proprietà mentale può essere realizzata da diverse proprietà fisiche, perciò identità di tipo mentale non implica identità di tipo fisico. Soprattutto quando si tratta di stati mentali come credenza e desiderio, l'obiezione della realizzabilità multipla appare ancora più convincente, poiché avere una certa credenza o un certo desiderio sembra poter essere realizzabile da più stati fisici. Pare un'assunzione molto vincolante che ogniqualvolta due soggetti hanno per esempio il desiderio di mangiare un gelato si trovino nello stesso identico stato cerebrale. Una risposta a questo tipo di obiezione consiste nel relativizzare l'identità alla specie, il che porterebbe i riduzionisti a sostenere che solo all'interno dei sistemi della stessa specie si può

[26] Gli stati intenzionali si distinguono perciò dai cosiddetti *qualia* (cui si accennerà più avanti nel testo), ovvero tutti gli stati dotati di qualità sensoriali, come i dolori, le sensazioni, le emozioni, ecc.

[27] Cfr. Putnam 1975, cap. 21 *The Nature of Mental States*.

[28] Ivi, p. 437.

parlare di identità tra stati mentali e fisici, o che comunque gli stati mentali a poter essere ridotti siano una classe ristretta e omogenea da rendere plausibile l'identità col tipo sottostante. Non dovremo identificare il dolore con una proprietà fisica P, ma il dolore negli umani con P, il dolore nei polpi con P*, e così via. Tuttavia, come già sottolineato in precedenza, stati mentali come la credenza o il desiderio non si prestano a questo tipo di riduzione.

Vi è poi un'ulteriore questione, quella relativa alle qualità soggettive delle esperienze coscienti (i *qualia*). Secondo molti autori, tra i quali il già citato Nagel, queste esperienze sarebbero irriducibili a proprietà fisiche poiché sono comprensibili solo da un punto di vista in prima persona (come per esempio il sapore di un gelato o il profumo di un fiore).

Un'altra critica importante arriva da Saul Kripke[29]. Egli afferma che, se anche supponessimo l'identità "mal di denti = avere le fibre C stimolate", questo tipo di identità sarebbe del tutto diversa rispetto a quella che vede "acqua = H_2O". Questo perché, in base alla sua teoria del riferimento dei nomi propri, che si basa sul concetto di *designazione rigida* (i designatori rigidi sono espressioni che identificano l'oggetto a cui si riferiscono, attraverso le sue proprietà essenziali, in tutti i mondi possibili), le identità come "acqua = H_2O" sono, se vere, vere *di necessità*, poiché "acqua" designa la stessa sostanza (in questo caso H_2O) in qualsiasi mondo possibile nel quale esiste; al contrario le identità come "mal di denti = avere le fibre C stimolate" sono contingenti, poiché è possibile che il dolore insorga per la stimolazione di altre fibre, o che nonostante le fibre C siano stimolate il soggetto non provi dolore; perciò quest'ultima identità non può essere vera *sempre*, in tutti i contesti e in tutte le occasioni, dunque non è necessaria. In conclusione, il mentale non può essere necessariamente fisico[30].

Un modo per ribattere a quanto afferma Kripke sarebbe quello di smentire il fatto che l'enunciato "mal di denti = fibre C stimolate" è contingente, mostrando che tale identità sembra contingente, ma è, in effetti, necessaria. Bisognerebbe cioè dimostrare che, quello che ci sembra, che percepiamo come dolore, non essendo realizzato cerebralmente tramite la stimolazione delle fibre C, in realtà non è dolore. Tuttavia, sottolinea Paternoster,

[29] Cfr. Kripke 1972.

[30] Simone Gozzano ritiene che la possibilità di provare dolore in assenza della stimolazione delle fibre C sia una mera illusione data da un'intuizione ingenua della nozione di sensazione (cfr. Gozzano 2007, pp. 116-42).

un'obiezione del genere appare poco plausibile, poiché è alquanto improbabile che ciò che percepiamo come dolore non sia dolore.

Alle obiezioni contro la teoria dell'identità dei tipi possiamo infine aggiungere che, oggigiorno, la neuroscienza fatica a offrire esempi concreti di identità tra stati mentali e stati cerebrali.

1.3.2 La teoria delle identità delle occorrenze

Anche a causa delle numerose critiche ricevute, la teoria dell'identità dei tipi è stata abbandonata e sostituita da una versione più moderata, la teoria dell'identità delle occorrenze, in base alla quale specifiche occorrenze di proprietà mentali sono identiche a specifiche occorrenze di proprietà fisiche, ovvero ogni evento mentale è identico a un evento cerebrale. Quindi ci sarebbe un unico evento, descrivibile però attraverso due vocabolari diversi, quello mentalistico e quello neurofisiologico.

Per chiarire meglio qual è la differenza delle due teorie e pertanto qual è la distinzione tra *occorrenze* e *tipi* facciamo un esempio: il mio mal di denti è un'occorrenza, il mal di denti in generale è un tipo.

La teoria dell'identità delle occorrenze risulta quindi meno vincolante rispetto a quella dell'identità di tipo, poiché identità di occorrenza non implica l'identità di tipo. Infatti, due soggetti diversi potrebbero trovarsi nello stesso stato mentale (per esempio credere che fuori splende il sole) e tuttavia trovarsi in differenti stati cerebrali, ovvero non necessariamente verranno attivati lo stesso gruppo di neuroni. Questo a dimostrare che allo stesso tipo di stato mentale (fuori splende il sole) non corrisponde sempre un particolare tipo di stato neuronale. Ed è proprio in tale modo che la teoria dell'identità delle occorrenze risulterebbe immune dall'obiezione della realizzabilità multipla, poiché ogni occorrenza, supponiamo di un dolore, è identica a un qualche evento fisico, benché non esista un singolo tipo di eventi fisici che sottostà a tutti i dolori.

Ciò che fondamentalmente la teoria dell'identità delle occorrenze afferma è che non esiste un unico stato cerebrale specifico che corrisponda ogni volta a uno stato mentale generale.

Questo tipo di teoria è alla base del monismo anomalo di Donald Davidson che analizzeremo di seguito.

1.4. Il monismo anomalo

Una versione della teoria dell'identità delle occorrenze è il monismo anomalo di Donald Davidson: "monismo" in quanto sostiene che gli eventi mentali sono eventi fisici; "anomalo" in quanto ritiene che gli eventi, quando sono descritti in termini psicologici, non ricadono sotto leggi rigorose. Ma vediamo come Davidson arriva a questa formulazione.

La teoria delle identità delle occorrenze serve a Davidson per conciliare la contraddizione, per lui solo apparente, di tre principi: il principio di interazione causale, il principio del carattere nomologico[31] della causalità e il principio dell'anomalia del mentale. Analizziamoli nel dettaglio.

Il *principio di interazione causale* sostiene che «almeno alcuni eventi mentali interagiscono causalmente con eventi fisici»[32]. Davidson identifica la percezione (dove le cause sono fisiche e gli eventi mentali) e l'azione intenzionale (dove invece le cause sono mentali e gli eventi sono fisici) come i casi dove l'interazione causale tra eventi mentali e fisici è più semplice da riscontrare, e precisa che l'interazione causale del mentale con il fisico riguarda tutti gli eventi mentali, se non in maniera diretta «magari mediante relazioni causali con altri eventi mentali»[33], ma nel caso si trovassero eventi mentali che non sono né cause né effetti di eventi fisici, allora il suo argomento non li riguarda.

Il secondo principio che Davidson prende in considerazione, come detto prima, è quello del *carattere nomologico della causalità*. Egli ritiene che dati due eventi qualsiasi in relazione di causa ed effetto, c'è sempre una legge sotto la quale essi possono essere sussunti: dove c'è causalità c'è una legge. Ogni relazione causale, in qualsiasi ambito, è esemplificazione di una legge della fisica.

Il terzo principio è quello *dell'anomalia del mentale*, ovvero il fatto che non ci sono leggi deterministiche in base alle quali poter prevedere e spiegare gli eventi mentali. Mentre i

31 Con "nomologico" si intende conforme a leggi di natura.
32 Davidson 1980, tr. it. p. 286.
33 *Ibid.*

processi fisici sono spiegabili tramite leggi "strette", cioè scientifiche, gli eventi mentali non sono spiegabili e prevedibili attraverso questo tipo di leggi.

Come ci spiega lo stesso Davidson, il paradosso che scaturisce dall'accettazione di questi tre principi riguarda il fatto che «i primi due principi […] insieme implicano che almeno alcuni eventi mentali possano essere previsti e spiegati in base a leggi, mentre il principio dell'anomalia del mentale nega proprio questo»[34]. Egli tuttavia ritiene che i tre principi siano tutti veri e dunque che non ci sia nessun paradosso, e intende dimostrarlo sostenendo il monismo anomalo. La coerenza dei principi verrà dimostrata da Davidson servendosi di una concezione identitaria del fisico e del mentale che non presenta alcuna contraddizione interna.

Davidson si oppone alle teorie presenti allora sulla relazione tra eventi mentali ed eventi fisici, che egli raggruppa in tre categorie: il *monismo nomologico*, secondo il quale il mentale è identico al fisico ed è sottoposto alle sue stesse leggi (per esempio la teoria dell'identità dei tipi di cui abbiamo parlato prima rientra in questo gruppo); il *dualismo nomologico*, dove egli classifica il parallelismo, l'interazionismo e l'epifenomenismo[35], e dove si afferma l'irriducibilità ontologica dei due ambiti, i quali tuttavia sono collegati da leggi psicofisiche; e infine il *dualismo anomalo*, che vede anch'esso il mentale ontologicamente irriducibile al fisico e inoltre nega leggi che mettano in relazione mentale e fisico.

L'ultima combinazione che resta, e che sarà quella scelta da Davidson, è il monismo anomalo, tesi che sostiene, come già detto, che tra mentale e fisico non vi sia distinzione ontologica, ma neppure correlazione nomologica:

> il monismo anomalo assomiglia al materialismo in quanto afferma che tutti gli eventi sono fisici, respingendo però la tesi […] che si possano dare spiegazioni puramente fisiche dei fenomeni mentali. Il monismo anomalo […] ammette la possibilità che non tutti gli eventi siano mentali pur continuando a insistere che tutti gli eventi sono fisici[36].

---

[34] Ivi, p. 287.

[35] Il parallelismo ritiene che il mentale e il fisico siano due regni senza alcuna relazione e tra i quali non c'è nessuna relazione causale. Per l'interazionismo il mentale e il fisico appartengono a due categorie ontologiche diverse, tuttavia tra di loro può esservi interazione reciproca. L'epifenomenismo infine sostiene che gli eventi mentali possono essere effetti di eventi cerebrali, ma non cause. Cfr. De Caro 1998, p. 101.

[36] Davidson 1980, tr. it. p. 293.

L'anomalia consiste proprio nel fatto che si presuppone l'assenza di leggi scientifiche che correlino i tipi di eventi mentali con i tipi di eventi fisici; tuttavia, questo non impedisce la corrispondenza tra singole occorrenze.

1.4.1. Gli eventi mentali

Gli eventi vengono definiti da Davidson come «individui irripetibili e collocati nel tempo»[37], sono delle occorrenze nello spazio e nel tempo che non necessariamente, come invece il termine evento potrebbe suggerire, si riferiscono a un cambiamento. Per esempio: «Tizio non si mosse davanti al pericolo» indica un evento, ma non un cambiamento. Per questa concezione un evento è insensibile al modo in cui viene descritto: la descrizione «il 20 luglio 1944 Klaus von Stauffenberg attentò a Hitler» e la descrizione «il 20 luglio 1944 un colonnello dell'esercito tedesco depose una valigetta contenente una bomba vicino al Führer» si riferiscono allo stesso evento solo se le due descrizioni possono essere inserite nello stesso spazio-tempo per l'evento descritto[38].

Il criterio che Davidson utilizza per definire il mentale non si basa sull'esser «privato, soggettivo o immateriale, bensì [sul] fatto d'esibire quel che Brentano chiamava intenzionalità»[39]. Egli tuttavia dà a questo criterio una caratterizzazione linguistica, asserendo che un certo evento è mentale se può essere descritto attraverso un verbo intenzionale, come credere, desiderare, ecc. Come ci fa notare De Caro, da questo tipo di concezione vengono lasciati fuori i cosiddetti *qualia*, ovvero gli stati qualitativi (classico esempio è quello del dolore). Egli ritiene che questa "assenza" sia dovuta al fatto che, entro la tradizione analitica, si ritenevano filosoficamente rilevanti solo gli stati intenzionali poiché connessi con il linguaggio[40]. Gli eventi mentali sono dunque quegli eventi nelle cui descrizioni occorrono uno o più predicati psicologici, dove appunto con "predicato psicologico" si intendono quei predicati che contengono un verbo che esprime un atteggiamento proposizionale, ovvero verbi quali credere, desiderare, sperare, ecc, (Tizio crede/desidera/spera che *p*).

---

37 Ivi, p. 288.
38 L'esempio è preso da Laudisa 1999, p. 22.
39 Davidson 1980, tr. it. p. 290.
40 Cfr. De Caro 1998, p. 103.

Un evento mentale è un evento descritto in un vocabolario mentale, viceversa un evento fisico è un evento descritto nel vocabolario della fisica; vale a dire che l'identificazione di evento mentale con evento fisico porta al fatto che ci sia un unico evento, il quale però ammette due descrizioni distinte: una in termini mentalistici e una in termini fisicalistici:

> ho supposto che di ogni evento psicologico particolare e datato possiamo dare una descrizione in termini puramente fisici; è così, per ogni classe d'eventi *data* e *finita*, possiamo stabilire una correlazione fra eventi psicologici e descrizioni fisiche[41].

Gli eventi mentali, descritti in termini psicologici, possono essere compresi, interpretati e anche previsti alla luce di principi normativi di razionalità e d'intelligibilità che non sono applicabili quando essi vengono descritti in termini fisici. L'attribuzione di eventi mentali è governata da principi di razionalità e di umanità, che ci portano a considerare le azioni degli altri come azioni compiute da persone, da creature razionali simili a noi in quanto dotate di coscienza e desideri:

> Non c'è modo d'attribuire credenze, una per una, a una certa persona sulla base del suo comportamento verbale, delle sue scelte o di altri segni locali, […] perché le credenze particolari risultano per noi sensate soltanto in quanto sono coerenti con altre credenze, preferenze, intenzioni, speranze, paure aspettative e così via. […] il fatto è che il contenuto di un atteggiamento proposizionale deriva dal posto che occupa nella struttura. […] Nella misura in cui non riusciamo a scoprire una struttura coerente e plausibile negli atteggiamenti e nelle azioni degli altri, stiamo rinunciando alla possibilità di trattarli come persone[42].

Importante è sottolineare il ruolo svolto dalle azioni per l'interpretazione: infatti, ogni atto interpretativo rimanda ai comportamenti e alle azioni di colui che vogliamo interpretare. È proprio partendo dalle azioni compiute dagli altri soggetti che, secondo Davidson, si può capire cosa gli altri intendono, e in questo modo attribuirgli stati mentali, ovvero interpretarli:

41 Davidson 1980, tr. it. p. 333.
42 Ivi, pp. 302-303.

> Quando domandiamo perché qualcuno ha agito così, vogliamo che ci si fornisca un'interpretazione. Il suo comportamento appare bizzarro, estraneo, eccentrico, senza scopo, non in carattere, sconnesso; o forse non riusciamo neppure a riconoscere in esso un'azione. Quando apprendiamo le sue ragioni, abbiamo un'interpretazione, una nuova descrizione di quel che ha fatto, descrizione che lo inserisce in un quadro familiare[43].

Quindi sostenere che un'azione è compiuta per una determinata ragione significa darne una spiegazione causale e significa anche mettere in risalto le credenze, i desideri e qualsiasi altro stato intenzionale del soggetto. Egli difende la natura causale della relazione tra ragione e azione, sostenendo l'identità di cause e ragioni e affermando che la ragione per un'azione ne è la causa. Spiegare un'azione equivale a fornire le ragioni per cui l'agente l'ha compiuta: ma se vogliamo fornire appunto una spiegazione dell'azione, dobbiamo chiederci quale relazione intercorre tra la ragione e l'azione, e tale relazione è per Davidson una relazione causale.

> Per la relazione tra una ragione e l'azione da essa spiegata, è fondamentale l'idea che l'agente abbia compiuto l'azione *perché* aveva quella ragione[44].

Davidson identifica nel desiderio e nella credenza la cosiddetta "ragione primaria": «fornire la ragione per cui un agente ha fatto qualcosa è questione di indicare l'atteggiamento favorevole […] o la credenza connessa […] o entrambe le cose; chiamerò questa coppia la ragione primaria per cui l'agente ha effettuato l'azione»[45].

Tuttavia lo stesso Davidson modificò parzialmente tale visione quando si rese conto che questa non prendeva in considerazione l'intenzione con cui viene svolta l'azione; inoltre pareva quasi che l'azione seguisse necessariamente dalla ragione primaria, quasi fosse una conclusione[46].

1.4.2. Le leggi psicofisiche

43 Ivi, p. 49.
44 *Ibid.*
45 Ivi, p. 42.
46 Cfr. De Caro 1998, pp. 84-89.

Davidson, nonostante ritenga che essere in uno stato mentale sia essere in un certo stato fisico, sostiene che i vocabolari che descrivono i due livelli non coincidono in quanto, come detto in precedenza, non esistono leggi psicologiche nello stesso senso in cui esistono leggi fisiche. Le regolarità del mondo fisico prendono forma di leggi che valgono in modo assoluto, privo di eccezioni, per questo vengono dette "strette". Le leggi del mentale invece sono leggi *caeteris paribus*, ossia «a parità di condizioni», a causa della varietà e quantità di circostanze che influenzano, o possono influenzare, la relazione che una tale legge esprime . L'anomalia alla quale si riferisce Davidson, come sottolineato in precedenza, risiede proprio in questo, cioè nel fatto che non possiamo correlare tramite leggi psicofisiche le descrizioni mentali con quelle fisiche, non ci sono esemplificazioni di leggi-ponte tra gli stati mentali e gli stati fisici, e questo avviene in base a una ragione di principio: la differenza tra le leggi della fisica e quelle di scienze come la psicologia è ontologica, «non c'è alcun senso importante in cui la psicologia possa esser ridotta alle scienze fisiche»[47].

Quindi, secondo Davidson, tradurre le spiegazioni psicologiche in spiegazioni fisicalistiche non è possibile, anche se ogni singolo evento mentale è effettivamente un evento fisico, perché non esiste alcuna legge psicofisica che correli eventi mentali tra loro simili a eventi fisici tra loro altrettanto simili, e tra l'altro «una conoscenza completa della fisica dell'uomo […] non produce necessariamente una conoscenza della psicologia»[48].

Questa incompatibilità delle proprietà mentali con le proprietà fisiche è data per Davidson dal carattere *olistico* e *normativo* del mentale. Ciò significa che non è possibile attribuire a una persona uno stato mentale senza attribuirgliene molti altri, poiché ogni stato mentale è strettamente connesso ad altri stati mentali. Per esempio, la mia credenza che Sassari sia a nord di Cagliari implica che io abbia la credenza che Sassari e Cagliari siano città, che il nord e il sud sono punti geografici di riferimento, e così via.

> Le credenze e i desideri si manifestano nel comportamento solo in quanto modificati e mediati da altre credenze, desideri, atteggiamenti, spostamenti

[47] Davidson 1980, tr. it. p. 345, Cfr. Gozzano 2007, pp. 16-20. Cfr. anche De Caro 1998, pp. 74-75.

[48] Davidson 1980, tr. it. p. 334.

> d'attenzione senza limite. È chiaro come questo olismo del regno mentale sia una spia dell'autonomia come pure del carattere anomalo del mentale[49].

Dunque è proprio la natura olistica della mente che impedisce che ci possano essere leggi psicofisiche.

Sebbene non esistano leggi strette che governano il rapporto fra credenze e desideri da un lato e azione dall'altro, la nozione di causalità, come abbiamo visto, trova applicazione nell'ambito dell'azione umana. Infatti, spiegare il comportamento di un individuo, ovvero trovare le cause, ci obbliga a interpretare le sue azioni come compiute da un soggetto razionale, dotato di desideri, credenze, che sono a noi comprensibili, ma che non sono riducibili a leggi universali (strette): «una comprensione completa del funzionamento del corpo e del cervello non rappresent[a] una conoscenza del pensiero e dell'azione»[50].

1.4.3. Sopravvenienza

La teoria proposta da Davidson determina un tipo di relazione tra mentale e fisico che va sotto il nome di "sopravvenienza". Questo concetto fu introdotto da George Moore[51], il quale si riferiva alla relazione che intercorre tra i predicati valutativi e quelli descrittivi: è impossibile per due soggetti concordare in tutti i predicati valutativi (buono, giusto) ma non in quelli descrittivi. Per esempio, se due persone hanno compiuto le stesse identiche azioni nelle stesse identiche circostanze (proprietà descrittive), non possono essere una buona e l'altra cattiva (proprietà valutativa). La bontà morale, anche se è irriducibile alle proprietà descrittive, è una proprietà che sopravviene sulle caratteristiche descrittive di una persona e sulle sue azioni. Quindi se c'è differenza dal punto di vista valutativo, ci deve essere anche dal punto di vista descrittivo, ma questo non implica che i predicati (o le proprietà) valutativi possano essere ridotti a quelli descrittivi[52].

La sopravvenienza dunque è una relazione di dipendenza piuttosto che di riduzione: essa comporta in maniera generale che una proprietà A sopravviene su una proprietà B nel caso in cui, se due oggetti sono diversi per una qualche proprietà A, allora lo sono anche per

[49] Ivi, p. 297.
[50] Ivi, p. 336.
[51] Moore 1912.
[52] Cfr. Davidson 1980, tr. it. p. 338; cfr. De Caro 1998, pp. 116-18; cfr. anche Nannini 2002, p. 126.

una qualche proprietà B, e tuttavia non è vero il contrario. Dunque le proprietà A sono sopravvenienti sulle proprietà B.

Davidson applica questa nozione al problema mente-corpo e sostiene al riguardo che:

> le caratteristiche mentali sono in qualche senso dipendenti da, o […] sopravvengono a caratteristiche fisiche. […] 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[53].

La nozione di sopravvenienza sostenuta da Davidson viene anche definita sopravvenienza "debole", poiché le proprietà mentali sopravvengono sulle proprietà fisiche se e solo se per ogni individuo *x* e ogni individuo *y*, se *x* e *y* non si distinguono per alcuna proprietà fisica, essi non si distinguono neanche per nessuna proprietà mentale.

Una versione "forte" della sopravvenienza per esempio intende quest'ultima come una relazione che, dal punto di vista ontologico, vale in tutti i mondi possibili. Davidson tuttavia, che in questo segue Quine, è molto cauto nell'impiegare nozioni modali, come quella di mondo possibile.

La sopravvenienza così come intesa da Davidson può essere vista come la sintesi di tre diverse idee relative al rapporto tra mentale e fisico, ovvero *covarianza*, *dipendenza* e *irriducibilità*[54].

La *covarianza* si riferisce al fatto che non possono esserci due stati identici sotto ogni aspetto fisico ma diversi per qualche aspetto mentale, quindi se due individui sono indiscernibili per quanto riguarda tutte le proprietà fisiche lo saranno anche riguardo alle proprietà mentali. Per *dipendenza* si intende che se non esistessero gli stati fisici non esisterebbero nemmeno gli stati mentali, poiché le caratteristiche dei secondi sono determinate da quelle dei primi. Infine con *irriducibilità* ci si riferisce al fatto che il mentale conserva una sua autonomia esplicativa in quanto una proprietà mentale non è riducibile a una proprietà fisica, non c'è per esempio una proprietà fisica con cui identificare – sempre e comunque – la credenza che Sassari è a nord di Cagliari.

---

53 Davidson 1980, tr. it. p. 294.

54 Cfr. Paternoster 2002, pp. 51-52; Da Caro 1998, pp. 117-18.

A questo punto Davidson ritiene di aver fornito tutti gli elementi per poter sostenere che il monismo anomalo concilia i tre principi esposti inizialmente. Bisogna dunque tener presente che «causalità ed identità sono relazioni tra eventi individuali, comunque descritti»[55]. Le leggi invece sono linguistiche e come tali spiegano o prevedono eventi assumendoli come esempi delle stesse, descrivendoli in un modo o nell'altro. Se si parla di accadimenti individuali non ha senso parlare di leggi che comportano la costanza e quindi la ripetibilità degli eventi. Il principio di interazione causale «tratta gli eventi in estensione e risulta cieco alla dicotomia fisico/mentale»[56]; il principio di anomalia del mentale riguarda invece «eventi descritti come mentali, giacché gli eventi sono mentali solo in quanto descritti come tali»[57]. Il carattere nomologico della causalità, così interpretato, non ci dice che «ogni asserzione singolare vera di causalità esemplifica una legge»[58], ma che «quando gli eventi stanno in relazione di causa ed effetto, possiedono descrizioni che esemplificano una legge»[59].

Numerose sono state le critiche mosse al monismo anomalo di Davidson e alla sua concezione di sopravvenienza, tra cui quella mossa da Jaegwon Kim il quale sostiene che la concezione di Davidson toglierebbe qualsiasi potere causale alle proprietà mentali, cadendo così nell'epifenomenismo. Ma di questo parleremo più avanti.

1.4.4. La critica di Jennifer Hornsby

In una serie di articoli[60] Jennifer Hornsby mette l'accento sulle difficoltà che il *principio del carattere nomologico della causalità*[61] pone al monismo anomalo di Davidson. La critica parte dall'analisi di un certo tipo di fisicalismo, di cui il monismo anomalo secondo l'autrice fa parte, e la concezione della causalità che da esso scaturisce.

La Hornsby inizia facendoci notare che, secondo il fisicalismo, ciò che distingue le varie scienze l'una dall'altra è l'utilizzo di differenti vocabolari, ovvero le diverse scienze, utilizzando ciascuna una propria tassonomia, ci forniscono diversi modi di classificare gli stessi eventi su un fondo comune di eventi fisici. A questo quadro va aggiunto il senso

55 Davidson 1980, tr. it. p. 294.
56 *Ibid.*
57 Ivi, p. 295.
58 *Ibid.*
59 *Ibid*.
60 Cfr. Hornsby 1980-1, 1985, 1999.
61 Lo ricordiamo, in base al principio del carattere nomologico della causalità dati due eventi qualsiasi in relazione di causa ed effetto, c'è sempre una legge sotto la quale essi sono sussunti.

comune che, pur non essendo una scienza, dà una propria classificazione degli stessi eventi descritti dalle altre scienze, ma in termini differenti. Dunque, se accettiamo il fisicalismo, possiamo dire che alla base ci sono gli eventi così come descritti dalla fisica, o meglio che esistono solo gli eventi descritti dalla fisica, e le altre scienze non fanno altro che ridescrivere gli stessi eventi in altri modi (e lo stesso fa il senso comune).

Gli eventi sono importanti, sottolinea la Hornsby, poiché le leggi della fisica parlano di eventi, le leggi della fisica possono spiegarci gli eventi, dunque possiamo dire che il mondo causale include gli eventi. Per questo parlare di eventi per la Hornsby è importante, poiché se riteniamo che alcuni eventi riguardano le persone, allora parlare di eventi ci consente anche di capire quale sia il posto degli esseri umani, e di conseguenza del mentale, nel mondo causale:

> physicalism, which is an ontological doctrine about the world's contents, leads naturally to a more substantial metaphysical doctrine about the world's workings, that under their physical descriptions, all events are susceptible of total explanations, of the kind paradigmatically afforded by physics in terms of physical laws and other physically described events[62].

Questo è il tipo di fisicalismo con il quale la Hornsby si vuole confrontare, fisicalismo che porta con sé l'idea che debbano esistere specifiche relazioni di identità tra gli eventi descritti in un certo vocabolario (poniamo quello mentale) e gli eventi descritti nel vocabolario fisico. Tale genere di fisicalismo dunque implica che «each mental event can be uniquely identified by a phrase making exclusive and essential use of predicates of some scientific theory (expression that are proprietary to neurophysiology, or biochemistry, say)»[63].

La Hornsby ritiene che il monismo anomalo di Davidson sia un caso particolare di fisicalismo inteso nei termini succitati. Davidson sostiene che gli eventi di cui parla la psicologia del senso comune (gli eventi mentali), sono gli stessi eventi di cui parlano le leggi della fisica (gli eventi fisici), e non è importante se la tassonomia del linguaggio della fisica e quella del senso comune non sono concordi. Infatti, è proprio a partire dall'assenza di relazioni strette tra questi due linguaggi che Davidson costruisce la sua tesi dell'identità tra gli eventi descritti nei due linguaggi. Il monismo anomalo così caratterizzato, ci fa notare la Hornsby, sembra del tutto coerente con il principio del carattere nomologico della causalità,

62 Hornsby 1985, p. 445.
63 Hornsby 1980-1, p. 73.

poiché una volta che gli eventi sono descritti fisicamente, questi possono essere spiegati nei termini delle leggi fisiche.

Tuttavia l'identità tra mentale e fisico affermata da Davidson, argomenta la Hornsby, presuppone una concezione, per lei erronea, che ogni tipo di fisicalismo a suo avviso accetta, ovvero la *concezione mereologica*: «the world of events seems always to have been viewed by philosophers mereologically – as exhausted by items that may be bigger or smaller, but whose identities depend upon nothing except their being of the right size»[64].

La mereologia può essere definita come «the theory of parthood relations: of the relations of part to whole and the relations of part to part within a whole»[65]; essa dunque caratterizza la logica delle relazioni tra la parte e il tutto, l'idea è che si costruirebbe il tutto (in questo caso i macro eventi mentali) attraverso la fusione (o la somma) delle parti di cui è composto (in questo caso i micro eventi fisici), se mettiamo assieme i micro eventi della fisica e li fondiamo, otteniamo gli eventi macro (come quelli del senso comune). La definizione stessa di mereologia introduce la relazione "è parte di": i micro eventi fisici sono "parti" di macro eventi mentali, i macro eventi sono visti come il risultato della fusione di micro eventi fisici. Le identità pertanto possono essere asserite tra eventi al livello macro (mentali) e fusioni di eventi al livello micro (fisico). In base a tale concezione quindi ogni entità è una parte spazio-temporale di un tutto che comprende, attraverso le fusioni, tutte le entità fisiche di base. La concezione mereologica presuppone che:

(A) *z* è la fusione di *x* e *y*.

Se il fisicalismo e il principio del carattere nomologico della causalità dipendono da una concezione mereologica degli eventi, screditando tale concezione degli eventi si getteranno ombre sia sul fisicalismo sia sul principio del carattere nomologico della causalità. Ciò che la Hornsby vuole dimostrare è che il principio del carattere nomologico della causalità e la concezione mereologica sono incompatibili, anche se il principio del carattere nomologico della causalità sembra dipendere dalla mereologia[66].

Proviamo a fare un esempio per capire meglio come funziona la concezione mereologica: si può pensare che un evento mentale, come per esempio la decisione di uscire

64 Ivi, p. 82.

65 Varzi 2011.

66 In realtà lo scopo della Hornsby è più ampio, nel senso che riguarda sia il fisicalismo sui *continuans* (sostanze persistenti nel tempo), sia il fisicalismo sugli eventi. Tuttavia per gli interessi della presente ricerca ci focalizzeremo solo sulla critica alla concezione mereologica legata al fisicalismo sugli eventi.

presa da Mary, sia identico a un evento fisico nel cervello di Mary, poiché riteniamo che ci sia una fusione di eventi microfisici che occupano la stessa regione spazio-temporale della decisione di Mary di uscire. Ma per la Hornsby questa caratterizzazione è problematica.

Dato che è nella natura degli eventi causare ed essere causati, prosegue l'autrice, ci aspettiamo che le fusioni di eventi dotati di potere causale abbiano a loro volta potere causale, dandoci delle descrizioni del perché una cosa segue a un'altra; invece sembra che in base alla concezione mereologica «fusions of things which cause and are caused need not themselves be things which cause and are caused»[67].

Parrebbe che, prosegue la Hornsby, se si importa la relazione "è parte di" dentro a un linguaggio scientifico (un microevento "è parte di" un macroevento), allora avremo tutte le risorse necessarie per costruire delle descrizioni che riguardano gli eventi di tutti i giorni (gli eventi mentali). Tuttavia, suggerisce l'autrice, non c'è ragione di pensare che le fusioni costruite da eventi usando la relazione di "è parte di" debbano essere a loro volta eventi: infatti, questi presunti eventi ai quali (A) ci impegna mancano di qualsiasi valore esplicativo. Questa concezione si porterebbe dietro l'idea che le relazioni parti-tutto non sono possibili solo dove è naturale pensare che sia così, «not only in the cases where there are special reasons for seeing events as comprised from other events», ma «much more generally»[68]. Dunque per la Hornsby non tutti i casi nei quali la mereologia viene applicata agli eventi sono problematici, in alcune circostanze la fusione è anche accettabile, ma ci sono casi in cui tale caratterizzazione è discutibile. Proviamo a pensare, prosegue la Hornsby, a eventi quali la morte di Giulio Cesare, la battaglia di Hastings e il discorso di Edward Heath. Questi, presi singolarmente, sono tutti degli eventi, ma uniti non fanno un unico evento, poiché l'evento che ne viene fuori non è causa di alcunché. Quali possono essere le cause e gli effetti di una tale fusione, si chiede la Hornsby, se le parti che occorrono in tale evento sono accadute rispettivamente nel 44 a.C., nel 1066 e nel 1984? L'evento che viene fuori da tale fusione non ha nessun potere causale. All'interno del dominio degli eventi (A) è incompatibile con ogni descrizione plausibile della relazione causale; la difficoltà con (A) consiste nel fatto che «it introduces fusions which, having no place in the causal nexus, lack any title to be identified with events»[69]. Questo problema non nasce solo per le fusioni tra eventi discontinui (come nell'esempio precedente), ma anche per eventi che hanno un legame spazio-temporale stretto,

67 Hornsby 1985, p. 450.
68 Ivi, p. 446.
69 Ivi, p. 452.

perché la sola contiguità spazio-temporale delle parti di queste fusioni non può essere ciò che è necessario per dotare tale elemento di uno statuto causale. L'evento che causa, per esempio, può essere fuori dallo spazio-tempo dell'evento effetto. Gli eventi che causarono, per esempio, il crollo della borsa di Wall Street il 24 ottobre del 1929 non si possono far rientrare nello spazio-tempo solo di quel 24 ottobre 1929.

Il verificarsi di un evento all'interno della regione spazio-temporale nella quale qualche altro evento occorre, da solo, non assicura che un evento sia parte dell'altro; dunque, secondo la Hornsby, per sostenere l'identità tra eventi mentali e eventi fisici non è sufficiente la concezione mereologica, che anzi ci dà una visione errata degli eventi. Di conseguenza, bisogna trovare qualche argomento per le identità che non assume semplicemente una concezione mereologica. Non si può dare per scontato che la fusione di due eventi sia necessariamente un evento: gli eventi descritti in un determinato vocabolario potrebbero non essere specificabili quando espressi in un altro linguaggio, nel senso che potrebbero mancare di una qualsiasi rilevanza esplicativa significativa.

Quanto detto finora, ci fa notare la Hornsby, ha grande rilevanza per la plausibilità del principio del carattere nomologico della causalità; infatti se gli eventi nel cervello sono legati l'un l'altro da leggi fisiche strette, allora siamo inclini a pensare a questi eventi nel cervello come a eventi microscopici, che possiamo individuare solo grazie alla neurofisiologia. Ma, si chiede la Hornsby, sono queste le stesse cose macro di cui parla la psicologia del senso comune? Se gli scienziati non fanno altro che studiare le cose micro, allora basta introdurre la relazione di "è parte di" per assicurare che gli eventi mentali possano essere inseriti tra le cose sussunte sotto leggi: «we may sometimes have to introduce the relation of parthood to ensure that mental events are among the things subsumed by laws»[70]. Sembra dunque che chi, come Davidson, pensa che tutti gli eventi relazionati causalmente possono essere sussunti in leggi fisiche, sostiene che gli eventi che riconosciamo sono in effetti fusioni di eventi descritti in base a leggi.

La visione che ne scaturisce per la Hornsby è che gli eventi mentali, che sono causalmente relazionati con altri eventi, siano troppo "grandi" per poter essere descritti nel vocabolario della fisica e quindi possano essere visti come composti di parti che possono essere descritte da leggi fisiche. Queste parti più piccole sono ovviamente gli eventi fisici i quali possono essere sussunti sotto leggi fisiche. Il ruolo causale del mentale verrebbe

70 Ivi, p. 457.

pertanto salvaguardato grazie a questa scomposizione dei macro eventi mentali in micro eventi fisici più piccoli.

Tuttavia, si chiede la Hornsby: «if the parts whose descriptions are found in laws are to do their job, then it has to be seen how they in turn are causally connected. […] In order to sustain the claimed connexion between causation and laws, a principle that entities built up from events subsumed by laws are (the same as) events subsumed by laws is required […]. Davidson's account of causation supports his anomalous monism, but is supported by the false constructionalist account of events»[71].

Quindi sembra che il principio del carattere nomologico della causalità implichi qualcosa come (A), ma abbiamo visto prima come per la Hornsby (A) è incompatibile con qualsiasi descrizione plausibile della relazione causale. Dunque il principio del carattere nomologico della causalità, che è un principio che descrive regolarità causali, è incompatibile con una concezione mereologica degli eventi, anche se esso dipende da una versione di fisicalismo che a sua volta dipende da tale concezione degli eventi.

La morale dunque, conclude la Hornsby, è che non è scontato che la fusione di due eventi sia anch'essa necessariamente un evento. Il monismo anomalo invece sembra assumere proprio questo, ovvero che la fusione di eventi fisici al livello micro – che è identica a un evento mentale particolare – è essa stessa un evento che può essere sussunto sotto una legge fisica, ma: «if events *e* and *f* are such as to need to be singled out by different ideologies, it will take an impressive argument to show that *e* is nonetheless the same as *f*»[72]. Anche includere la relazione di "è parte di" nel vocabolario della fisica non è sufficiente, conclude la Hornsby.

Sebbene Davidson non usa la concezione mereologica come fanno certi fisicalisti per sostenere l'identità tra eventi mentali ed eventi fisici (l'identità sostenuta da Davidson, lo ricordiamo, è tra occorrenze e non tipi), il principio del carattere nomologico della causalità comporta per l'autrice che «a fusion of physicist's events is the same as some mental event»[73].

Per la Hornsby c'è un'oggettiva difficoltà nell'affermare in che modo il fisico fissa il mentale, o come il mentale dipende dal fisico, «and certainly no contribution is made (the task

71 Hornsby 1980-1, pp. 87-88.
72 Hornsby 1985, p. 454.
73 Ivi, p. 458.

is only made to seem as if it were more smooth or simple than it actually is) by presupposing that a common ontology has to underlie dependencies between the two realm»[74].

1.4.4.1. Il problema della transitività

Un altro problema che la Hornsby mette in luce parlando del principio del carattere nomologico della causalità è quello della *transitività*. Come sappiamo, in base al principio del carattere nomologico della causalità, ogni volta che due eventi sono relazionati causalmente c'è una legge che li sussume. Tuttavia per l'autrice questa caratterizzazione renderebbe problematica la spiegazione delle relazioni transitive, cioè di relazioni dove *a* causa *b* e *b* causa *c* e in virtù di ciò *a* causa *c*. Se riteniamo che ci sia una legge che sussume *a* e *b* e un'altra che sussume *b* e *c*, non abbiamo ragione di pensare, argomenta l'autrice, che ve ne sia una che sussume anche *a* e *c*[75]:

> We might call this the problem of transitivity. More strictly it should be called a problem of non-intransitivity, because it will afflict Davidson's account so long as the causal relation is not intransitive, and it is allowed that *a* might cause *c* "via" its causing *b*[76].

Un modo per ovviare a questo problema sarebbe quello di ricorrere alla concezione mereologica degli eventi, ma abbiamo visto che per la Hornsby il principio del carattere nomologico della causalità è incompatibile con la concezione mereologica: «earlier, a principle about the properties of fusions […] was found to be implausible. We see now that the same principle […] would be needed to justify the identity claims that certain causal claims impose»[77].

Ogni volta che siamo in presenza di un evento macro, come per esempio una valanga, per poterlo spiegare in termini di leggi strette, dobbiamo vederlo come composto da tanti piccoli eventi fisici che sono relazionati causalmente l'uno con l'altro, relazioni che possono essere sussunte sotto leggi fisiche appunto. Grazie alle relazioni transitive, gli effetti al livello

74 Hornsby 1980-1, p. 89.
75 Cfr. ivi, p. 87.
76 *Ibid*.
77 Ivi, p. 88.

macro della valanga possono essere spiegati facendo riferimento alle cause e agli effetti degli eventi al livello micro. In questo modo si potrebbe spiegare come le connessioni causali al livello micro causano le connessioni causali al livello macro, ma «in the presence of intransitivity, there seems to be no way in which one might seek to show how the causation of rough-hewn everyday events was grounded in connexions between their law-governed components»[78].

La transitività dunque pone un serio problema al principio del carattere nomologico della causalità, poiché per poter rendere conto delle relazioni transitive tale principio deve ricorrere alla mereologia ma, come abbiamo detto, c'è incompatibilità tra la mereologia e il principio del carattere nomologico della causalità. Il problema della transitività mette ulteriormente in luce, pertanto, come il principio del carattere nomologico della causalità è al contempo dipendente e incompatibile con la concezione mereologica degli eventi, quindi non può essere un principio che fornisce una descrizione soddisfacente della causalità.

1.4.4.2. Le azioni

La Hornsby nega pertanto l'identità tra eventi mentali ed eventi fisici così come caratterizzata dal fisicalismo in generale e da Davidson in particolare, supportando tale affermazione anche con considerazioni sul ruolo causale del mentale nello spiegare le azioni, azioni che abbiamo visto hanno uno statuto rilevante nel monismo anomalo.

Anche in questo caso l'autrice continua a porre l'accento sul problema della causalità, sostenendo che quando si parla di azione spesso si confonde tra le azioni e le loro descrizioni. Davidson mantiene questa distinzione poiché gli permetterebbe di mantenere la nozione di agente: «to have exhibited agency, a person has to have done something intentionally»[79]. Questo perché per Davidson un'azione è un evento che è intenzionale in base a una certa descrizione. Tuttavia non bisogna confondere, sottolinea la Hornsby, la parola "azione" così come usata dal senso comune e la caratterizzazione che ne viene data in filosofia. Le azioni del senso comune sono ciò che le persone fanno, le azioni in filosofia si riferiscono a cose nel dominio degli eventi, le azioni in filosofia sono eventi e indicano elementi all'interno di un'ontologia di cose particolari. Per esempio, argomenta la Hornsby, «Anna's writing of the

78 Ivi, pp. 87-88.
79 Hornsby 1999, p. 625.

world 'blue', Peter's setting light to the fuse, […] in ordinary English […] are not actions. For in ordinary English "actions" is used for things that people do. And a person does not relate to what she does as she relates to her action (in the philosophers' sense just introduced): people do not do events»[80]. Per il linguaggio comune ciò che Anna fa secondo la Hornsby è, per esempio, «*[to] write the world 'blue'*»[81].

In base all'argomentazione della Hornsby, anziché sostenere, come fa Davidson, che possiamo avere diverse descrizioni di ciò che una persona fa, sarebbe meglio dire che abbiamo modi diversi di descrivere cose diverse fatte da quella persona. Dunque l'importanza che Davidson dà alla distinzione tra azione e descrizione dell'azione stessa può essere caratterizzata diversamente: dobbiamo concedere che ci possono essere due cose diverse che un individuo fa e che il fatto che quell'individuo fa una delle due è la stessa cosa di quando fa l'altra; tuttavia una delle due cose potrebbe essere stata fatta intenzionalmente e l'altra no. Infatti, le cose che le persone fanno hanno varie descrizioni, le quali possono essere (o meno) intenzionali: «when it is recognized that a single event may be described as a person's doing one thing and described as her doing some other thing […], then it must be allowed that she may have done one thing intentionally, another not»[82].

Facciamo un esempio: se sulla via per Tebe Edipo discute con un uomo e lo uccide, senza sapere che si tratta di suo padre, allora sotto la descrizione "uccidere un uomo" l'azione di Edipo è intenzionale, ma sotto la descrizione "uccidere suo padre" la sua azione non è intenzionale[83]. In questo caso quindi non diciamo che l'uccisione da parte di Edipo di un uomo e l'uccisione di Edipo del padre sono due azioni; diremo piuttosto che la sua azione, il suo fare, è ridescrivibile in due modi diversi, per esempio in questo caso come omicidio e come patricidio.

Le cose che le persone fanno, argomenta l'autrice, non sono cose ridescrivibili, ma sono cose ripetibili (chiunque può uccidere un uomo, o il proprio padre). Affermando che sono ridescrivibili le stiamo confondendo con le azioni (che appunto sono particolari ridescrivibili). Le azioni invece sono il fare più cose contemporaneamente, una cosa che fa Epido è uccidere un uomo, un'altra cosa che fa è uccidere suo padre: «his action is a doing of

80 Ivi, p. 624.
81 *Ibid*.
82 Ivi, p. 625.
83 L'esempio è ripreso da Friggieri 1991, pp. 18-9.

(at least) these two kinds»[84]. Parlare di descrizioni può reintrodurre la confusione che la distinzione tra cose che le persone fanno e azioni è intenzionata a eliminare: «of course things that people do are *not* redescribable things, they are repeatable things; when spoken of as if they were redescribable, they get confused with actions (which *are* redescribable particulars)»[85].

Secondo la Hornsby è un errore pensare che per ogni azione ci sia un desiderio, una credenza sottostante che viene citata ogni volta che si richiede una spiegazione di tale azione. Questo errore è dovuto proprio al fatto che si usa il termine "azioni" a volte per gli eventi, altre volte per le cose che fanno le persone, e quindi ci si dimentica che ciò che viene spiegato non è un evento, ma «an agent's doing something that she did»[86].

Nel monismo anomalo c'è una difficoltà, argomenta la Hornsby, nel trovare una descrizione di relazione causale che concili il fatto che alcuni eventi mentali interagiscono con eventi fisici (principio di interazione causale) con il principio nomologico della causalità, soprattutto se consideriamo che ci sono cose che le persone fanno intenzionalmente che non sembrano essere eventi, come per esempio non andare intenzionalmente a un incontro. Oppure che ci può essere una spiegazione causale di un'azione, dove il fatto che il soggetto che la compie non crede qualcosa è fondamentale; in questo c'è qualcosa che può essere sussunto sotto una legge, si chiede la Hornsby. La ragione non può essere la causa dell'azione poiché «action concepts do not provide us with anything exact to say in the language of neurophysiology»[87].

Il problema dell'argomento di Davidson, sostiene la Hornsby, è che il nostro uso del linguaggio del mentale non ci fornisce nessun elemento che possa andar bene per essere identificato con ciò che viene riportato nelle leggi fisiche: «we can ask at what stage in the neural chain we find the man's action. Where is the action itself, as distinct from its effects, as distinct from the events that the agent brought about?»[88].

Supponiamo, prosegue la Hornsby, che Mary crede che può andare dall'altro lato della strada attraversandola, e il desiderio di Mary di andare dall'altro lato della strada causa il suo attraversamento della strada. Un ostacolo nel trattare "Mary crede che … e il desiderio di

84 Ivi, p. 19.
85 Hornsby 1999, p. 625.
86 Ivi, p. 628.
87 Hornsby 1980-1, p. 81.
88 Ivi, p. 77.

Mary di … causa una determinata azione" deriva dalla presenza della "e" in ciò che precede la causa:

> Should we say that wherever *x* and *y* caused *a*, there is something *x-and-y* which caused *a*? Or should we say that wherever *x* and *y* caused *a*, each of *x* and *y* caused *a*? It seems that unless we say one or other of these things, we shall not have the kind of causal statement which asserts a relation between two items, and which can thus be brought into such a relation to a statement of law that nomological vocabulary can be seen as applicable to the same things as mental vocabulary applies to[89].

Potrebbe sembrare che le ragioni, quando sono causa dell'azione, possano essere gli elementi mentali da identificare con il fisico, ma non è così: «reasons are in fact quite the wrong sort of things to be causes»[90]. Quando si tratta di trovare una spiegazione razionale, ciò che ritroviamo nel mondo causale non sono le ragioni, bensì le persone, le quali hanno ragioni (*reasons*). Quindi, conclude la Hornsby, se vogliamo trovare un dominio comune che fornisca la verità sia per le frasi che contengono vocaboli mentali, sia per le frasi che contengono vocaboli fisici, dobbiamo volgere l'attenzione alle persone: «if, then, we want to show, as Davidson wanted to, that a common domain provides for the truth both of statements containing everyday mental vocabulary and of statements containing physical vocabulary, we must advert to persons (who have reasons)»[91].

Davidson risponde alle critiche della Hornsby affermando che è possibile difendere l'affermazione che le azioni sono sempre degli eventi: infatti le azioni riflettono scelte tra le diverse opzioni che una persona ha, e lo scegliere di fare una cosa anziché un'altra è un modo differente di descrivere un'azione: «there are endless causes for any event»[92].

Per Davidson la Hornsby non centra il punto del suo monismo anomalo quando si chiede come sia possibile descrivere uno stato mentale, o un'azione, in termini fisici che siano adatti per le leggi strette, le leggi della fisica. È parte del perché parlare in termini psicologici è irriducibile ai termini fisici, continua Davidson, che non c'è speranza di una corrispondenza

89 Hornsby 1999, p. 630.
90 Ivi, p. 632.
91 *Ibid*.
92 Cfr. Davidson, 1999, p. 639.

in modo sistematico tra elementi di un vocabolario con elementi di un altro vocabolario: se questo fosse possibile, la psicologia non sarebbe irriducibile alla fisica.

Nonostante ciò, nonostante appunto come afferma Davidson «I have no hope that anyone will discover how to describe a state of mind or an action in physical terms that are suitable for strict laws»[93], la critica della Hornsby mette in luce la difficoltà del monismo anomalo di spiegare, proprio come dice il titolo di un suo articolo "Which Physical Events are Mental Events?"[94].

La Hornsby quindi pone l'accento sulle problematiche poste dalla caratterizzazione della causalità del monismo anomalo; questo ci pare rilevante, tenuto anche in considerazione che, come vedremo, a considerazioni simili arriva la critica di Kim. La Hornsby e Kim infatti, seppur utilizzando argomentazioni diverse, mettono in evidenza il fatto che non si capisce quale sia la relazione tra la causa fisica e la causa mentale nel monismo anomalo: «it now seems that everyday psychology does not deliver causal statements which can be brought into the kind of relation with statements of law»[95].

Il nocciolo della critica sia della Horsnby sia, come vedremo, di Kim, pare dunque essere lo stesso: il problema della causalità del mentale. Se non c'è connessione causale coperta da leggi tra eventi mentali ed eventi fisici, sembra che gli eventi non possano essere legati tra loro.

1.5. Il funzionalismo

Il funzionalismo è stata una delle posizioni in filosofia della mente di maggior successo nella seconda metà del secolo scorso, e che tutt'oggi, seppur con sensibili modifiche, continua ad avere dei sostenitori. Questo grazie al fatto che offre una prospettiva che, come vedremo dettagliatamente più avanti, si discosta sia dal dualismo che dal materialismo tradizionale, in quanto ritiene che esista un senso nel quale il mentale è diverso dal fisico, anche se presumibilmente non esistono sostanze non fisiche, e ogni occorrenza mentale è in sé solamente fisica, tuttavia la caratterizzazione del mentale non corrisponde ad una descrizione fisica (per esempio, la proprietà di avere un dolore non è solo la proprietà di avere certi

93 *Ibid*.
94 Cfr. Hornsby 1980-1.
95 Hornsby 1999, p. 632.

neuroni stimolati, come invece abbiamo visto riteneva il riduzionismo. Come vedremo infatti, il funzionalismo individua gli stati mentali in modo autonomo rispetto alla loro realizzazione fisica, cerebrale: per il funzionalismo gli stati mentali sono stati computazionali. Secondo la tesi funzionalista, uno stato mentale è ciò che è in virtù del fatto che svolge una certa funzione, ovvero per il fatto di ricoprire un certo ruolo all'interno dell'intera attività mentale di un soggetto. Lo stato mentale in sostanza corrisponderebbe allo stato funzionale.

Per i funzionalisti il funzionamento di una mente umana può essere paragonato al funzionamento di una macchina di Turing; la quale è il modello di qualsiasi calcolatore digitale. Questa macchina è composta da una testina di lettura e un nastro potenzialmente infinito diviso in caselle, sulle quali viene stampato un simbolo. Mentre la macchina nel suo complesso si trova in un certo stato, la testina può eseguire un certo numero di operazioni, come spostarsi a sinistra o a destra, leggere l'eventuale simbolo, cancellarlo, sostituirlo, e fermarsi. Le varie operazioni definiscono la "tavola di transizione". L'identità di uno stato della macchina dipende dalle relazioni che esso ha con gli input, gli output, e gli altri stati del programma, e due stati sono identici se assolvono alle medesime funzioni nelle medesime condizioni di input, output, e relazioni con altri stati. Non importa con quali materiali si realizza la macchina, ma quanti stati può avere e che relazioni possono stabilirsi tra essi: la macchina è dotata di un insieme finito di stati logici, ma tale sua descrizione logica non comporta nulla riguardo la natura fisica di questi stati, e neanche della macchina. Una volta che la macchina viene realizzata, ha in linea di principio un'infinità di stati strutturali.

La tesi dei funzionalisti consiste nel fatto che l'intelligenza può essere realizzata da adeguati programmi per computer: l'importante è il programma, ossia il procedimento usato per eseguire una certa funzione, e non il supporto fisico. Il fondatore di questa corrente di pensiero, e cioè Hilary Putnam[96], sostiene che uno stato mentale è analogo a uno stato della macchina di Turing, perciò la mente può essere studiata in modo del tutto indipendente dal suo sostrato fisico. Vediamo nel dettaglio di cosa si tratta.

1.5.1. Il funzionalismo di Hilary Putnam

[96] Cfr. Putnam 1975, cap 16 *Brains and Behavior*; cap.18 *Minds and Machine*; cap. 20 *The Mental Life of Some Machines*; cap. 21 *The nature of mental states*.

Intorno agli anni ’60 Hilary Putnam propone una teoria che riguarda il mentale, la quale prenderà il nome di “funzionalismo”; questa è basata sul fatto che gli eventi mentali non sono costituiti dalle loro proprietà intrinseche, ma dalle relazioni causali che essi intrattengono con altri eventi mentali, con gli input sensoriali e con gli output comportamentali: gli eventi mentali sono cioè caratterizzati e individuati dal loro ruolo funzionale. Gli stati mentali sono identificabili con stati funzionali; facendo un esempio, si può dire che esiste qualcosa che è in comune a ciascun tipo di dolore e che fa sì che sia un dolore: questo qualcosa è un particolare stato funzionale dell’organismo.

Per dimostrare questo Putnam propone un’analogia tra la mente umana e i programmi delle macchine calcolatrici, nello specifico, come abbiamo detto, una macchina di Turing. L’attenzione è dunque rivolta, più che alle parti fisiche del calcolatore, ai programmi: la mente viene paragonata a un *software* come il corpo a un *hardware*. La macchina di Turing, insomma, viene vista come un modello per la mente, e attraverso questo tipo di analogia si pensava di poter trovare risposta ai secolari quesiti sulla natura e sulle proprietà degli stati mentali.

Tale analogia può essere spiegata attraverso il concetto di *isomorfismo funzionale* tra due sistemi, una relazione che lega tutti quei sistemi per i quali vi è «a corrispondence between the states of one and the states of the other that preserves functional relations»[97]. Da ciò segue che «two systems can have quite different constitutions and be functionally isomorphic»[98]. Per esempio, un calcolatore elettronico può essere isomorfo a un uomo che usa carta e penna. nonostante lo stato chimico e fisico sia differente, la «functional description may be the same»[99].

Tutto questo porta Putnam a concludere che una spiegazione in termini di fisica o chimica del funzionamento della nostra mente sarebbe inutile, poiché non sarebbe in grado di dirci come funziona la mente: i problemi della vita mentale non riguardano il livello materiale (fisico-chimico). Infatti,

97 Ivi, cap. 14 *Philosophy and our mental life*, p. 291.
98 Ivi, p. 292.
99 Ivi, p. 293.

> what we are interested in is: can we say[…] since we have this sorto f structure, this sorto f program, it follows that we will be able to learn this, we will tend to like that, and so on[100].

Questo implica che gli stati mentali possono essere studiati indipendentemente dal modo in cui sono concretamente realizzati, estendendo così la possibilità di forme di attività mentale in entità non umane, come gli animali e i computer. Nulla vieterebbe di ammettere che stati mentali identici, cioè che conducono allo stesso tipo di effetti, possano essere realizzati in maniera differente: così si possono studiare in modo comparato processi mentali in specie differenti, negli uomini e nelle macchine, trasferendo le conoscenze ottenute in un dominio agli altri.

Ciò che portò Putnam al funzionalismo, come egli stesso afferma, fu il desiderio di difendere il materialismo, mostrando come questo non implichi il riduzionismo che, come abbiamo visto, sosteneva la riduzione delle nozioni del mentale a quelle fisiche o, come afferma Putnam, l'idea che la «nostra materia sia più importante della nostra funzione, che il nostro *cosa* sia più importante del nostro *come*»[101], sostanzialmente che l'essenza delle nostre menti sia costituita dal nostro *hardware*.

Per questo inizialmente il funzionalismo di Putnam si centrò sul problema mente-corpo proponendo, come già detto, l'analogia tra la mente umana e una macchina di Turing; egli infatti ritiene che la macchina di Turing e un essere umano, se descritti con un certo livello di astrazione, potrebbero funzionare più o meno nello stesso modo:

> una macchina, un essere umano, una creatura dotata di una chimica al silicone e uno spirito disincarnato potrebbero funzionare grosso modo nella stessa maniera se descritti a un adeguato livello d'astrazione[102].

Putnam ritiene infatti che il problema mente-corpo non abbia niente a che vedere con l'esperienza umana soggettiva, ma che piuttosto sorge in «connetion with any computing system capable of answering questions about its own structure»[103]. Questo perché, sempre

100 Ivi, p. 298.
101 Putnam 1988, tr. it. p. 11.
102 *Ibid*.
103 Putnam 1975, cap. 18 *Minds and Machines*, p. 362..

secondo Putnam, il comportamento di una macchina calcolatrice non è spiegato dalla fisica o dalla chimica della macchina, ma è spiegato dal programma della macchina, il quale, benché realizzato in una particolare fisica o chimica, è comunque «an abstract property of the machine»[104]. Analogamente, le proprietà psicologiche degli esseri umani non sono proprietà fisiche e chimiche, anche se da queste possono essere analizzate, divenendo così anch'esse una proprietà astratta, proprio come il programma delle macchine. Dunque, per il livello di astrazione che presenta la psicologia, una mente può essere identificata con una macchina di Turing, o meglio *è* una macchina di Turing. Non ci sarebbero differenze tra la psicologia della nostra specie e la struttura di una macchina di Turing: per questo gli stati mentali vengono descritti e studiati nel linguaggio delle macchine di Turing, che è un linguaggio logico-matematico, invece che in quelli della psicologia del senso comune, basata su credenze, sentimenti, passioni. In questo modo non ci sarebbe bisogno di nessuna ipotesi sull'esistenza e sulla natura della mente, né di alcuna sua riduzione ad altra sostanza.

Il funzionalismo evidenzia quindi quali operazioni svolge la mente, non quali sostanze (fisico-chimiche, spirituali, ecc.) sono impiegate per la realizzazione di tali operazioni; per questo differenti stati fisici possono realizzare il medesimo stato mentale.

Solo successivamente, e a partire proprio dalle riflessioni fatte finora, Putnam si interrogò sulla natura degli stati mentali. Anche qui il punto di partenza è l'accostamento tra macchina di Turing e esseri umani.

Dunque l'identificazione non è più solo tra la mente e la macchina, ma più direttamente tra gli stati mentali e gli stati della macchina: per ogni stato mentale esiste uno stato di macchina, funzionalmente equivalente allo stato mentale; per determinare questa equivalenza funzionale egli sottolinea la necessità che «we should know of physical state on the basis of which we could not merely *predict* human behaviour, but causally explain it»[105].

Qualsiasi difficoltà sorga per chi voglia identificare, come fanno i materialisti, gli stati mentali con degli stati cerebrali si ripresenterebbe anche per l'identificazione tra gli stati funzionali della macchina e i suoi stati fisici. Se si trova strano dire che un dolore è un'attività delle fibre C, allora sarà ugualmente strano affermare che la macchina si trova in un certo stato funzionale perché, ad esempio, il suo flip-flop 36 è aperto. Ma, dato che questa affermazione non è affatto strana, anzi ci stupiremo se la macchina non funzionasse come

[104] Ivi, *Introduction*, p. xiii.
[105] Ivi, cap. 18 *Minds and Machines*, p. 380.

funziona in base a certi processi fisici che avvengono nel suo interno, questo, sottolinea Putnam, non dovrebbe essere un problema neanche quando si tratta della relazione tra stati mentali e stati cerebrali. Nessuno ritiene un mistero il rapporto tra gli stati logici e gli stati strutturali di una macchina, e quindi in base all'analogia da lui proposta non lo sarà neanche interrogarsi sulla natura del rapporto tra stati mentali e stati fisici di un essere umano.

E anche se, in seguito a un accrescimento delle nostre conoscenze sulla natura fisica e sulle cause del comportamento degli esseri umani, enunciati che egli ritiene oggi devianti, come "il dolore è identico alla stimolazione delle fibre C", diventassero non devianti, ossia acquistassero un uso nel linguaggio, questo non dimostrerebbe che "avere le fibre C stimolate" e "provare dolore" sono lo stesso stato. Infatti, anche se enunciati che prima erano devianti possono acquisire un uso, questo non avviene perché le parole di cui sono composti gli enunciati acquistano significati nuovi, bensì perché i vecchi significati determinano un nuovo uso, dato dal nuovo contesto[106].

> The moral, I believe, is quite clear: it is no longer possible to believe that the mind-body problems a genuine theoretical problem, or that a "solution" t it would shed the slightest light on the world in wich we live. […] if the so-called "mind-body problem" is nothing but a different realization of the same set of logical and linguistic issues, the nit must be just as empty and just as verbal [107].

Dunque per Putnam ciò che ci interessa è la forma, non la materia, perciò la domanda che ci si dovrebbe porre non è "siamo fatti di materia o spirito?", bensì "qual è la nostra forma intellettuale?"[108]. E questo perché la sostanza, cioè ciò di cui siamo fatti, non stabilisce nessuna restrizione alla nostra forma. Putnam sostiene l'autonomia del "programma" dalla sua realizzazione materiale. La descrizione della struttura fisica che realizza i processi di pensiero non fa parte della loro essenza.

Uno stato mentale è dunque

---

106 Putnam al riguardo si rifà a delle osservazioni di Paul Ziff sul significato, secondo il quale si dovrebbe guardare alla teoria linguistica nel suo aspetto diacronico, ossia studiare i cambiamenti linguistici nel corso del tempo. Ziff sostiene che il significato è un'entità astratta, non riducibile alla concretezza dei singoli impieghi già dati, soggetta a stipulazioni continue che ne ampliano le possibilità di applicazione sotto la spinta di fattori sociali circoscritti e variabili nel tempo (cfr. Ziff 1966).

107 Putnam 1975, cap. 18 *Minds and Machines*, p.384.

108 Cfr. ivi, p. 330.

another *kind* of state, […] a functional state of a whole organism[109].

Esattamente come una macchina di Turing può essere descritta al livello delle istruzioni, così i processi mentali umani possono essere descritti al livello logico, come sequenze di stati mentali governate da leggi che ne controllano l'ordine; e così come una stessa macchina di Turing può avere realizzazioni fisiche del tutto diverse (quindi l'identità della macchina è indipendente dal suo supporto materiale), anche l'identità dei processi mentali umani è indipendente dal loro supporto materiale (il cervello).

1.5.2. Funzionalismo e riduzionismo

Abbiamo visto che, secondo il funzionalismo di Putnam, il ruolo funzionale di uno stato mentale coincide in sostanza con il ruolo causale dello stato, ossia con le relazioni che esso ha con input, output e altri stati funzionali. È questa l'argomentazione da preferire, sostiene Putnam, soprattutto se confrontata con altre tesi come il riduzionismo e il comportamentismo. Come accennato, non ci occuperemo del comportamentismo e prenderemo in considerazione solo la critica al riduzionismo.

Abbiamo già parlato dell'obiezione della "realizzabilità multipla" sostenuta da Putnam contro il riduzionismo, e in particolare contro la teoria dell'identità di tipo la quale, lo ricordiamo, sosteneva che tipi di stati mentali, come un dolore o una credenza, sono identificabili con tipi di stati cerebrali, come l'attivazione di certe fibre. Si era anche visto che in questo modo sembrava che qualora ci trovassimo in uno stato come il dolore, condivideremo lo stesso tipo di stato cerebrale di un individuo di qualsiasi altra specie quando prova dolore. Ed è su questo punto che, abbiamo visto, si centra la critica di Putnam. Egli sostiene che la sua ipotesi è la migliore perché identifica gli stati mentali in base alla loro organizzazione funzionale. Ciò che ci si aspetta è che in specie diverse i vari stati psicologici, o almeno i fondamentali, abbiamo tra loro e con il comportamento quelle che Putnam chiama

[109] Ivi, cap. 21 *The nature of mental states*, p. 433.

"probabilità di transizione" più o meno simili, ovvero ci si aspetta che, per esempio, un essere umano e un qualsiasi animale assettato assumano lo stesso comportamento, e cioè cercare da bere. Per Putnam qualsiasi organismo che si ritiene capace di assumere determinati stati, sembrerà possedere un certo tipo primitivo di organizzazione funzionale. Specie biologiche diverse infatti possono condividere le proprietà mentali, perché il modo in cui queste sono fisicamente implementate non incide sul loro ruolo funzionale, analogamente a quanto accade per i programmi delle macchine.

Dunque, prosegue Putnam, sarebbe molto più semplice sviluppare un programma che scopra leggi psicologiche che non siano specie-specifiche, ossia che possano riguardare specie diverse, il quale «porterà con sé un abbozzo del genere di organizzazione funzionale necessaria e sufficiente per un certo stato psicologico, oltre che una definizione precisa della nozione di "stato psicologico"»[110], piuttosto che un programma che sviluppi le leggi neurofisiologiche indipendenti dalla specie, come sostenevano i riduzionisti.

Si potrebbe anche affermare che il funzionalismo contenga in qualche modo una versione di riduzionismo, in quanto riduce gli stati mentali a strutture input-output. Ma, come fa notare Ned Block, il funzionalismo fornisce una riduzione senza eliminare gli stati mentali stessi: infatti, mentre il riduzionismo, affermando che gli stati mentali sono stati fisici eliminava di fatto gli stati mentali, il funzionalismo, presentando una particolare caratterizzazione degli stati mentali, si impegna nei confronti della loro esistenza[111]. Egli tuttavia sottolinea che è possibile trovare un senso in cui uno stato funzionale si può dire sia fisico. Per esempio, uno stato funzionale può essere considerato fisico se ogni sistema che ha quel determinato stato funzionale è un oggetto fisico, oppure se ogni realizzazione di quello stato funzionale (cioè ogni stato che gioca il ruolo causale specificato dallo stato funzionale in questione) è uno stato fisico[112], ma questo non implica un'identificazione come quella proposta dal riduzionismo.

1.5.3. Distacco dal funzionalismo

110 Ivi, pp. 469-70.
111 Cfr. Block 1980, pp. 176-77.
112 Cfr. ivi, p. 178.

Intorno agli anni ’70 Putnam tuttavia inizia a prendere le distanze da quanto visto finora, egli ritiene che il suo funzionalismo si basi su assunzioni troppo riduzionistiche, poiché, anche se non si riducono gli stati mentali direttamente a stati fisici, essi sono pur sempre riducibili a stati di macchina. Vediamo più dettagliatamente come arriva a questo distacco.

Verso la fine degli anni ‘60[113] Putnam perfezionò il concetto di macchina di Turing, così come illustrato precedentemente, introducendo quello di automa probabilistico, ovvero un dispositivo simile alla macchina di Turing «except that the transitions between “states” are allowed to be with various probabilities rather than being “deterministic”»[114], e le cui capacità di memoria hanno un limite fisso finito[115]. Dunque queste sono macchine fornite di organi di senso mediante i quali possono esaminare l’ambiente in cui si trovano e di organi motori che esse sono in grado di regolare.

L’ipotesi che gli stati mentali fossero stati funzionali veniva ulteriormente avvalorata, secondo Putnam, dal fatto che:

(1) All organisms capable of feeling pain are Probabilistic Automata.
(2) Every organism capable of feeling pain possesses at least one Description of a certain kind (i.e. being capable of feeling pain is possessing an appropriate kind of Functional Organization).
(3) No organism capable of feeling pain possesses a decomposition into parts which separately possess Descriptions of the kind referred to in (2).
(4) For every Description of the kind referred to in (2), there exists a subset of the sensory inputs such that an organism with that Description is in pain when and only when some of its sensory inputs are in that subset .
(5) For every Description of the kind referred to in (2), there exists a subset of the sensory inputs such that an organism with that Description is in pain when and only when some of its sensory inputs are in that subset[116].

[113] Putnam 1975, cap. 20 *The Mental Life of some Machines*, p. 408; cap. 21 *The Nature of Mental States,* p. 429.
[114] Ivi, cap. 21 *The Nature of Mental States*, p. 433.
[115] La macchina di Turing altro non è che un tipo speciale di automa probabilistico, le cui probabilità di transizione da uno stato all’altro sono 0,1.
[116] Putanm 1975, cap. 21 *The Nature of Mental States*, p. 434. Qui Putnam parla dei dolori, ma il discorso può essere generalizzato a qualsiasi stato mentale.

Un automa può essere realizzato da infinite configurazioni fisiche, di cui il cervello umano è solo un esempio. Una descrizione della psicologia con un vocabolario puramente fisico non offrirebbe nessuna spiegazione dell'organizzazione funzionale, ma darebbe una descrizione arbitraria di una delle macchine che realizzano l'automa che ci interessa.

Egli era consapevole che questa ipotesi fosse vaga, tuttavia la riteneva la migliore nel panorama filosofico del tempo, perché si prestava «to investigation of both a mathematical and an empirical kind»[117]. Questo avrebbe permesso alla psicologia, il cui unico scopo per Putnam era quello di produrre modelli meccanici di organismi, di fornire una forma normale della descrizione psicologica di organismi arbitrari, e non solo degli esseri umani. Una volta che la psicologia avrà fatto progressi nella ricerca saremo in grado di confermare l'ipotesi che gli stati mentali sono stati funzionali nello stesso modo in cui confermiamo le ipotesi della fisica:

> The laws of unreduced psychology, to the extent that they are true, will be explained by the fact that the psychological states they speak of are really these functional states, just as the laws of unreduced optics, to the extent that they were true, were explained by the fact that the "light rays" and "light waves" they spoke of were really electromagnetic radiation of certain wavelengths[118].

Individuare uno stato di una macchina di Turing vuol dire fornire le condizioni necessarie e sufficienti per essere quello stato. Egli riteneva che una teoria psicologica esauriente dovesse mettere in evidenza «illuminating differences between any possible psychological theory of a human being, or even functional description of a computing machine […] and a physical or chemical description»[119]. E la differenza più importante è che i sistemi che sono modelli della stessa teoria psicologica, sistemi che sono psicologicamente o funzionalmente isomorfi, non devono essere nello stesso stato fisico per trovarsi nello stesso stato mentale[120].

Però anche se consideriamo gli esseri umani come queste particolari macchine di Turing, o automi probabilistici, questi devono essere comunque in grado di avere stati analoghi a quelli delle macchine di Turing, e affinché questo avvenga tali stati devono essere

117 Ivi, p. 435, Cfr. Anche Guttenpaln 1994, p. 509.
118 Guttenplan 1994, p.509.
119 Putnam 1975, cap. 14, *Philosophy and our Mental Life,* p. 292.
120 Cfr. Guttenplan 1994, p. 510.

stati in cui gli esseri umani possono trovarsi «at any time, indipendently of learning and memory; and be total instantaneous states of the human beings – states which determine, together with learning and memory, what the next state will be, as well as totally specifying the present condition of the human being»[121].

Ma, fa notare Putnam, in base a queste caratteristiche nessuno stato psicologico può essere uno stato di una macchina di Turing. Anche se gli esseri umani fossero macchine di Turing, i loro stati di macchine non sarebbero identici ai loro stati psicologici. Questo perché le descrizioni offerte di macchina di Turing e essere umano sono a due livelli di organizzazione completamente diversi: per ogni essere umano esistono potenzialmente infiniti stati psicologici, mentre, per definizione, gli stati di una macchina sono finiti. Dunque è impossibile identificare gli stati mentali con quelli di una macchina di Turing.

E se nessun singolo stato di macchina può essere uno stato psicologico, tanto meno può esserlo una disgiunzione di stati di macchina. Per mostrare questo, argomenta Putnam, basta pensare a stati come la gelosia, l'amore, ecc., ossia a stati che non sono istantanei, ma comportano, per esempio, informazioni sulle relazioni sociali tra le persone. Ma se gli stati delle macchine di Turing sono istantanei e indipendenti dall'apprendimento e la memoria, uno stato come la gelosia o l'amore non può essere uno stato di macchina né una disgiunzione di stato di macchina.

Come egli stesso afferma, qui sostituisce la nozione di macchina di Turing con la nozione «of the sort of description that will be provided by an ideal psychological theory»[122].

Successivamente però Putnam si trova ad ammettere che la teoria psicologica che aveva immaginato con il suo funzionalismo, ovvero una teoria che fornisce una completa descrizione dei nostri stati psicologici così come una tavola della macchina di Turing fornisce una descrizione completa di uno stato funzionale di un computer, è una completa utopia. Infatti, mentre l'organizzazione funzionale di una macchina di Turing o di un automa probabilistico è data dalla tavola di transizione della macchina, la descrizione funzionale di un essere umano «might well be something quite different and more complicated»[123].

Questa utopia si fa ancora più grande una volta che Putnam pone l'accento anche sull'ambiente: il contenuto di una credenza non può essere un'entità solamente mentale, in

[121] Putnam 1975, cap. 14, *Philosophy and our Mental Life*, p. 298.
[122] Guttenplan 1994, p. 510.
[123] Putnam 1975, cap. 20, *The mental life of some machines*, p. 424.

quanto è possibile, a parità di stati cerebrali, intrattenere credenze diverse, aventi cioè diverso contenuto, in contesti ambientali differenti[124]. Tra l'altro egli sostiene che:

> io non credo neppure che tutti gli esseri *umani* che hanno lo stesso pensiero (in diverse culture, o con differenti basi di conoscenza e diverse risorse concettuali) abbiano in comune una caratteristica fisica e computazionale che possa essere "identificata con" quel pensiero. Il "livello intenzionale" non è riducibile al "livello computazionale"[125].

Quindi ora egli ritiene che il contenuto dei nostri pensieri non è determinato semplicemente dalla nostra organizzazione funzionale (che, lo ricordiamo, è solamente una questione di input sensoriali, transizioni da uno stato all'altro e output comportamentali), ma è determinato dalle nostre relazioni con le cose, l'ambiente e le altre persone.

Dunque Putnam abbandona l'idea di ridurre l'intenzionalità, ovvero gli stati mentali quali credenza e desiderio, a qualcosa come le relazioni funzionali.

1.5.4. Il funzionalismo liberalizzato

Tuttavia Putnam crede ancora in una forma di funzionalismo – in una versione "liberalizzata" di funzionalismo, come lui stesso la concepisce – che identifica «mental activities and capacities with ways of functioning, and that is compatible with the idea that those ways of functioning can, in principle, be realized in different physical systems, but minus the "science fiction"»[126]. La "science fiction" si riferisce alla convinzione di poter identificare gli stati mentali con gli stati funzionali, senza tener nel dovuto conto l'esternismo e l'anti-individualismo semantico. Il funzionalismo dell'odierno Putnam è quello che trova ancora le sue radici in Aristotele, il quale vedeva il mentale (la *psyche*) non come un organo o una cosa materiale o immateriale, ma come un insieme di capacità, funzioni e modi, in diretto contatto col mondo, in cui gli esseri umani sono organizzati per funzionare: «in that wide sense of the term, I am still a functionalist»[127]. L'essere umano è concepito come un

[124] Cfr. Putnam 1988, tr. it. p. 101.
[125] Putnam 1987, tr. it. pp. 26-27.
[126] Putnam 2012, cap. 3 *Corresponding with Reality*, p. 74.
[127] Putnam 1997a, p. 32.

organismo le cui funzioni sono "transazionali", cioè coinvolgono l'ambiente dall'inizio. Putnam dunque crede ancora nell'idea che ciò che conta per le proprietà mentali è la giusta capacità funzionale, ma queste funzioni non sono esclusivamente interne e non sono descrivibili nel vocabolario della scienza o in qualsiasi modo fissato in anticipo[128].

Ciò consente quindi che «Commander Data and I could both have the same mental states (a corollary of my original version of functionalism that I still subscribe to)»[129]; tuttavia, precisa Putnam, questo discorso non vale per i *qualia*, per gli stati qualitativi. Essi non sono identici a stati (o eventi) funzionali, bensì a stati (o eventi) neurologici, e questo porterebbe alla conseguenza che, sebbene un computer e un umano possono avere le stesse credenze, paure, ecc., «the question whether any relation of "identity" exists between the qualia of nonconspecifics (such as Commander Data and myself) is problematic. I would say, "Probably not"»[130].

1.5.5. Un'evoluzione del funzionalismo: la TCRM

Un esponente importante del funzionalismo è Jerry Fodor, il quale ha formulato la cosiddetta teoria computazional-rappresentazionale della mente (TCRM). Questa nasce dall'unione della teoria rappresentazionale della mente (TRM) e la teoria computazionale della mente (TCM). La prima analizza il rapporto tra rappresentazione e rappresentato, la seconda invece afferma che i processi mentali hanno accesso soltanto alle proprietà formali, e non semantiche, delle rappresentazioni. Fodor perciò sostiene la tesi che i processi cognitivi sono computazioni su rappresentazioni mentali. Gli stati mentali avrebbero natura rappresentazionale, ossia veicolerebbero informazioni sul mondo, rappresenterebbero il mondo in un determinato modo, perciò sarebbero valutabili semanticamente, cioè suscettibili di essere veri o falsi. Per esempio la credenza che Sassari è a nord di Cagliari è vera perché c'è nel mondo un fatto che la rende vera.

Le rappresentazioni mentali di cui parla Fodor sarebbero di natura linguistica poiché:

1. le parti costituenti si combinerebbero fra loro in base alle regole di una sintassi;

---

128 Putnam 2012, cap. 3 *Corresponding with Reality* , p. 83.

129 Ivi, cap.36 *How to be a Sophisticated "Naïve Realist"*, p. 637. Il Comandante Data è un personaggio della serie televisiva di fantascienza *Star Trek,* è una forma di vita artificiale ed è progettato per sembrare umano essendo dotato di una forma avanzata di intelligenza artificiale.

130 *Ibid*.

2. i costituenti che hanno una struttura semplice sono portatori di un significato che può essere un individuo o una proprietà del mondo;
3. sono composizionali, cioè il loro significato viene determinato dal significato dei loro costituenti e dalla loro struttura sintattica;
4. le espressioni di tipo enunciativo hanno un valore di verità e intrattengono tra loro relazioni logiche di implicazione.

Le ultima tre proprietà si riferiscono a questioni semantiche, un aspetto assai discusso in filosofia della mente.

Il problema consiste proprio nel fatto che, essendo le regole computazionali puramente formali, sintattiche, ovvero legate alla formulazione di formule, non si capisce come sia possibile che abbiano una valenza semantica, ossia come possano portare alla comprensione del significato. Fodor identifica la soluzione nelle regole logiche, le quali sono valide indipendentemente dal significato attribuito ai simboli che mettono in relazione. Espressioni che seguono regole logiche sono quelle del tipo «Nessun uomo non sposato è sposato», dove il valore di verità dell'enunciato non cambia per ogni possibile interpretazione, e quindi significato, che viene attribuita ai termini "uomo" e "sposato".

Dunque per Fodor «la forma logica di un pensiero sopravviene alla forma sintattica della rappresentazione mentale corrispondente»[131].

Per quanto riguarda invece gli aspetti computazionali si può dire che siano sensibili alle proprietà formali delle espressioni o ai simboli del linguaggio del pensiero, ma non al significato, per questo si può avere un'analogia con la macchina di Turing. Infatti il processo di lettura dei simboli in una macchina di Turing non ha niente a che fare con il significato di tali simboli. I processi mentali sarebbero sensibili solamente alle proprietà fisiche dei simboli che usano, e questo perché le espressioni del linguaggio e del pensiero sono pure iscrizioni materiali.

I processi mentali sono «sequenze causali di stati mentali»[132], questi vengono identificati come «computazioni, ossia operazioni definite sulla sintassi delle rappresentazioni mentali, e conservano attendibilmente la verità in un numero illimitato di casi»[133]. Per esempio la computazione di un essere umano, inizierà con la successione logica di stati di credenza o pensiero, ossia stati mentali, che consentiranno al soggetto di arrivare a una certa conclusione

131 Fodor 2000, tr. it. p. 25.
132 Ferretti 2001, p. 125.
133 Fodor 2000, tr. it. p. 25.

partendo da certe premesse, o di produrre un certo output comportamentale. Perciò ogni volta che un soggetto si trova in un certo stato mentale, egli intrattiene una relazione con un simbolo del linguaggio del pensiero (o «mentalese»). È importante sottolineare come una stessa iscrizione nel linguaggio del pensiero può avere un ruolo funzionale differente. Facciamo un esempio: se desidero che fuori splenda il sole esisterà un corrispettivo simbolo in mentalese, raffigurabile come «fuori splende il sole» in una fantomatica scatola del desiderio. Analogamente esisterà una fantomatica scatola delle credenze col simbolo «fuori splende il sole», qualora creda che fuori splende il sole. Perciò la stessa iscrizione in mentalese «fuori splende il sole» ricopre due ruoli funzionali differenti, nel primo caso quello di desiderio, nel secondo quello di credenza.

Tutto questo per mostrare che gli stati mentali ereditano dalle espressioni del linguaggio del pensiero con cui sono in relazione le loro proprietà sintattiche e semantiche. Il ruolo causale dello stato mentale, che è appunto il principio stesso di identità dello stato, è determinato dalle proprietà sintattiche dell'espressione con cui lo stato si trova in relazione. Il contenuto semantico, ossia ciò che rende vera o falsa la mia credenza che fuori splende il sole, è costituito dalle proprietà semantiche dell'espressione in mentalese «fuori splende il sole», vale a dire dalle sue condizioni di verità, ovvero il modo in cui il mondo deve essere affinché l'espressione in questione sia vera o falsa.

Il fatto che il ruolo causale venga determinato dalla sintassi è dovuto al fatto che un certo stato mentale ne causa un altro, dipendendo tutto questo solamente da regole legate alle proprietà formali delle iscrizioni, allo stesso modo in cui in una macchina di Turing un certo stato è causato dallo stato precedente, poiché nella macchina è codificata una certa istruzione che dice alla macchina stessa cosa deve fare in quella situazione.

Si potrà quindi ora capire la forte analogia tra la mente e il software di un computer: le connessioni causali tra stati mentali sono realizzate grazie alla regole sintattiche che governano la manipolazione delle rappresentazioni, ossia la sintassi del linguaggio del pensiero.

Riassumendo, per la TCRM gli stati mentali, oltre a essere analoghi agli stati di una macchina di Turing, ricoprono un ruolo formale determinato dalle relazioni sintattiche esistenti tra i simboli del linguaggio del pensiero. Essere in un certo stato mentale equivale a intrattenere una relazione con un simbolo del linguaggio del pensiero. Questo simbolo ha proprietà sia causali che semantiche: quelle causali sono determinate dalla sua forma, ovvero

dalle caratteristiche materiali del simbolo stesso; le proprietà semantiche invece derivano dal fatto che il simbolo rappresenta qualcosa, che ne determina l'essere vero o falso. Il ruolo principale è perciò svolto dalla sintassi, infatti «le proprietà causali di un simbolo vengono connesse alle proprietà semantiche attraverso la sintassi»[134]. Gli stati mentali ricevono perciò dalle espressioni con cui sono in relazione sia le proprietà sintattico-causali sia quelle semantiche, ed è grazie al fatto di essere in relazione con i simboli di un linguaggio che gli stati mentali hanno tali proprietà. Analogamente una macchina di Turing è concepita in modo tale che

> trasformerà un simbolo in un altro se, e solo se, i simboli così trasformati stanno in certe relazioni semantiche. […] I computer sono una soluzione al problema della mediazione tra le proprietà causali dei simboli e le loro proprietà semantiche[135].

1.5.6. Critiche al funzionalismo

Una delle maggiori critiche mosse al funzionalismo è quella secondo cui esso non è in grado di dar conto degli stati *qualitativi*, cioè le esperienze pure, come per esempio le emozioni e i dolori, o ogni vissuto fenomenologico soggettivo che accompagna molti dei nostri stati mentali (i *qualia* di cui abbiamo parlato in precedenza). Non terrebbe conto, in sostanza, dell'aspetto intrinseco degli stati qualitativi, di quell'aspetto cioè che è indipendente dalle relazioni con gli altri stati. In altre parole, in uno stato mentale sarebbe presente un aspetto che nessuna relazione esterna è in grado di individuare: di conseguenza, non può essere tradotto in termini funzionali. Ciò vuol dire che gli aspetti qualitativi possono essere colti solamente dal soggetto che ne fa esperienza, poiché l'unico modo per sapere cosa un soggetto sta provando è essere quel soggetto.

Questa per esempio è la critica di Thomas Nagel[136], della cui posizione dualista abbiamo parlato in precedenza. Egli sostiene che il funzionalismo ignora il "lato soggettivo", il quale non trova riscontro al livello degli stati delle macchine. Ciò che Nagel sostiene è che egli non rifiuta il fatto che si possa dare una caratterizzazione funzionale degli stati mentali,

134 Ferretti 2001, p. 128.
135 Ivi, p. 129.
136 Cfr. Nagel 1991.

piuttosto nega «che con l'aver stabilito una cosa del genere la loro analisi debba considerarsi conclusa»[137].

Egli, lo ricordiamo, ritiene che molti stati e processi mentali possono essere conosciuti soltanto in prima persona, ovvero solo essendo il soggetto nel quale si svolgono quei processi. Questa obiezione colpisce il fatto, sostenuto dal funzionalismo, che i processi mentali siano passibili di conoscenza oggettiva o, come si usa dire, "in terza persona". Nell'esperimento di Nagel, per comprendere la fisiologia di un pipistrello non abbiamo bisogno di condividere il sistema percettivo, ma per sapere che effetto fa essere un pipistrello sì. Anche se una macchina è in grado di riprodurre i processi mentali, questa non potrà riprodurre cosa si prova ad avere quel determinato stato mentale.

A questa critica i funzionalisti replicano affermando che non si preoccupano di questi aspetti perché essi non sono un oggetto di conoscenza: non c'è conoscenza possibile dei fenomeni in prima persona, ciò che si prova non sarebbe una "cosa". Spesso si ribatte a questa replica dei funzionalisti che essa non soddisferebbe un filosofo, in quanto un'indagine sugli stati mentali non può escludere e prescindere dal suo aspetto più importante, vale a dire la soggettività, l'io che prova determinate cose.

Un altro tipo di critica arriva da Ned Block il quale afferma che il funzionalismo classifica come menti dei sistemi che non lo sono. Si immagini, dice Block, che tutta la popolazione della Cina sia impegnata a realizzare un programma che descrive uno stato mentale di un essere umano, diciamo il mal di denti di Tizio. Ogni cinese realizzerà un passo di tale programma e quindi la descrizione funzionale della Cina sarà identica a quella di Tizio quando ha il mal di denti. Ma la Cina non proverà un mal di denti. Anche in questo caso, dunque, il funzionalismo fallisce nel cogliere l'aspetto caratteristico di quello stato mentale: la qualità del dolore. Ma non solo questo: infatti il requisito dell'identità di organizzazione sarebbe troppo debole e generico per qualificare una mente, ossia un sistema che dovrebbe agire, percepire, ragionare, ecc. La sola organizzazione pare quindi una proprietà troppo astratta per spiegare completamente la natura della mente. William Lycan ha cercato di aggirare questa critica affermando che solo nel caso in cui gli stati svolgano una funzione che soddisfi alcuni scopi del sistema complessivo, solo in questo caso una certa organizzazione funzionale può essere considerata una mente.

[137] Nagel 1974, tr. it. p. 380.

Alcune critiche che vengono mosse alla TCRM si basano sull'idea che una proprietà fondamentale degli stati mentali, la loro semanticità, ossia il loro riferirsi a oggetti e situazioni del mondo reale, non può essere catturata da un sistema formale, quale invece è la mente per il funzionalismo computazionale. Paternoster individua una delle più frequenti critiche al funzionalismo computazionale in quella che afferma che le teorie come la TCRM sono vincolate alla cosiddetta "condizione di formalità", cioè all'idea che i processi cognitivi siano sensibili esclusivamente alle proprietà non semantiche delle rappresentazioni. Ma non sarebbe questo il caso degli esseri umani: infatti l'introspezione e il comportamento mostrerebbero chiaramente come i nostri processi cognitivi siano "dipendenti dal significato". Gli esseri umani sono agenti intenzionali guidati da stati rappresentazionali[138].

Un esempio pratico di questo tipo di critica è l'argomento della *stanza cinese* di John Searle[139], con il quale egli pretende mostrare che qualunque sistema artificiale non potrà mai produrre e decodificare significati perché, anche se fornisse prestazioni che sembrano semantiche, esso sarebbe semplicemente programmato (computazionalmente) per eseguirle, quindi non le eseguirebbe al modo "intenzionale" delle persone. L'esperimento consiste nell'avere all'interno di una stanza un individuo che non conosce il cinese ma è dotato di un manuale, un libro di regole nella sua madrelingua che gli permette di rispondere correttamente in cinese alle domande che gli vengono formulate, sempre in cinese, dal di fuori della stanza. Egli associa semplicemente le giuste risposte alle domande in base alle corrispondenze dettate dal manuale e in questa sua attività di manipolazione di meri simboli formali non ha modo di incrementare minimamente la sua conoscenza del cinese.

Ciò che Searle vuole mostrare con questo esperimento è che se l'uomo all'interno della stanza non capisce il cinese, nonostante sia fornito del manuale di regole (analogo di un programma da calcolatore) per capire il cinese, neanche un calcolatore potrà capire il cinese, perché non ha accesso alle proprietà semantiche dei simboli che manipola. Così come non è l'attività dell'uomo nella stanza a conferire significato ai simboli, analogamente non sono le operazioni formali di un programma ad assegnare significato ai simboli che manipola. Searle afferma che se allo stesso uomo che si trova dentro la stanza gli facessero delle domande nella sua madrelingua, poniamo l'italiano, il modo di rispondere a queste domande sarebbe completamente diverso rispetto a quelle in cinese. Nel caso del cinese infatti le risposte erano

138 Cfr. Paternoster 2002, p. 41.
139 Searle 1980.

dovute alla manipolazione di simboli formali non interpretati, e non alla comprensione di tali simboli, come invece avverrebbe nel caso dell'italiano. Pertanto, un programma di calcolatore non sarà sufficiente per la comprensione, come invece sostengono i sostenitori del funzionalismo computazionale.

Risposte a questa critica arrivano da Douglas Hofstadter[140] e William Rapaport[141]. Hofstadter contesta a Searle che il quadro che egli propone, ossia l'uomo che esegue manualmente tutte le operazioni al pari di un calcolatore, è fondamentalmente sbagliato poiché riflette una «concezione assolutamente non realistica del rapporto tra intenzionalità e manipolazione di simboli»[142]. Searle ometterebbe dettagli fondamentali, come quello per esempio che per un essere umano sarebbe complicatissimo e richiederebbe un lavoro lunghissimo simulare manualmente un programma da calcolatore, ed è praticamente impossibile il fatto che egli riesca a incorporare tutto il manuale di regole e renderle reperibili ogni qualvolta ne ha bisogno.

La comprensione si trova tutta nel libro di regole e quasi niente nell'essere umano. Perciò il fatto che il programma in questo caso sia rappresentato da un individuo, e dunque da un essere dotato di intenzionalità, per Hofstadter non avrebbe alcuna importanza. Egli sostiene che in realtà Searle prende in esame un solo punto di vista, ovvero quello dell'essere umano, trascurando il secondo punto di vista, ossia quello del sistema.

William Rapaport ritiene invece che non sia così impossibile, come Searle afferma, far scaturire la semantica dalla sola sintassi. Rapaport[143] afferma che la sintassi di per se stessa è sufficiente per la semantica: la comprensione della semantica è la comprensione della sintassi. Dunque qualsiasi agente cognitivo, compreso un computer, capace di produrre sintassi, ovvero di manipolare simboli, sarà in grado di comprendere il linguaggio. Egli sostiene infatti che un agente cognitivo A con competenze riguardo al linguaggio naturale capisce gli output del linguaggio naturale di un altro agente cognitivo, B, costruendo e manipolando i simboli di un modello interno (un'interpretazione) degli output di B, considerati come un sistema formale. Il modello interno di A sarebbe una rappresentazione conoscitiva e un sistema di

140 Hofstadter e Dennett 1981.
141 Rapaport 1995.
142 Hofstadter e Dennett 1981, tr. it. p. 361.
143 Cfr. Rapaport 1995, pp. 49-88.

ragionamento che manipola simboli. Dunque A comprende la semantica di B tramite un'operazione sintattica[144].

Paternoster individua proprio nello scarso successo dei risultati conseguiti dall'intelligenza artificiale nell'elaborazione del linguaggio naturale un punto di forza dell'obiezione di Searle.

Una decina di anni dopo aver proposto l'argomento della stanza cinese Searle proporrà un'altra argomentazione basata sul fatto che

> la nozione di sintassi è essenzialmente relativa a un osservatore. Il fatto che processi computazionalmente equivalenti possano realizzarsi attraverso una molteplicità di supporti differenti non significa semplicemente che quei processi siano astratti: significa piuttosto che, dipendendo da un'interpretazione esterna, essi non sono affatto intrinseci al sistema[145].

Egli sostiene infatti che «gli stati computazionali non possono essere *scoperti* all'interno di un sistema fisico, ma vengono *assegnati* al sistema»[146]. Questo tipo di argomentazione è differente da quello della stanza cinese: con quell'esperimento Searle pretendeva mostrare che la semantica non è intrinseca alla sintassi; con l'argomentazione proposta ora, invece, egli mette in luce come la sintassi non sia intrinseca alla fisica. Infatti per caratterizzare sintatticamente un calcolatore è necessaria la presenza di un osservatore esterno «che assegni una interpretazione sintattica alle proprietà puramente fisiche del sistema»[147].

Un'ulteriore critica a questa visione viene da Paul Churchland[148], il quale afferma che ci sono davvero tipi e classi di computer che non ce la fanno a "pensare" e a "capire". Ma se si prendessero in considerazione dei "computer paralleli", ossia molto simili nel funzionamento al cervello umano, il quale viene definito da Churchland "macchina parallela", nel senso che i segnali in entrata, i cosiddetti input, vengono elaborati simultaneamente in miliardi di canali diversi. Quindi questi computer non si limitano a manipolare simboli riferendosi a regole immagazzinate in memoria, ma appunto riproducono il modo di elaborare

144 Cfr. ivi, p. 51.
145 Searle 1992, tr. it. p. 225.
146 *Ibid*.
147 Ivi, p. 226.
148 Churchland 1995, tr. it. pp. 262-63.

i dati del cervello. Secondo Churchland quindi se prendissimo in considerazione i computer paralleli ciò che sostiene Searle non avrebbe senso. Infatti per Churchland non esisterebbe qualcosa come "il significato intrinseco", cioè un significato disconnesso e indipendente dalle relazioni con gli altri stati e con il resto del mondo. Se così fosse, avverte Churchland, dovrebbe essere possibile che un sistema fisico possieda un solo stato interno, uno stato dotato di significato intrinseco e tutti gli altri stati dovrebbero essere privi di significato. Egli difende una posizione che va sotto il nome di *connessionismo*, il quale si fonda sull'idea che i processi cognitivi sono particolari tipi di computazioni eseguite da "reti neurali", ossia da strutture dinamiche costituite da un certo numero di unità interconnesse, ognuna delle quali simula, con un certo livello di astrazione, il comportamento di un neurone o di un complesso di neuroni reali[149]. Il connessionismo dei Churchland è perciò il cosiddetto connessionismo PDP (*parallel distributed processing*), fondato su una tesi "emergentista" secondo cui il pensiero è una proprietà emergente dall'autorganizzazione di una mente concepita come rete di unità profondamente interconnesse in modo distribuito[150].

Un'altra critica al funzionalismo arriva da Hubert Dreyfus[151], il quale sostiene che esso appare dualista perché non tiene conto che le capacità cognitive sono radicate sia dentro un corpo, il quale possiede determinate abilità percettive, motorie e sociali, sia all'interno di un contesto storico e culturale, con dei fini di natura sociale e comunicativa.

Dreyfus perciò ritiene che svincolare, come farebbe il funzionalismo, la competenza linguistica e cognitiva dal corpo significa presupporre che fra questi non vi sia alcun legame. Il limite delle macchine non sarebbe dovuto, per Dreyfus, al fatto che esse non hanno una mente, bensì al fatto che non hanno un corpo. Se le capacità umane saranno eguagliate da un computer, si baseranno sul seguire delle regole; i computer non possono far altro. Un computer ha bisogno di regole perché gli oggetti con cui ha a che fare devono essere definiti chiaramente e quello che deve essere fatto in seguito deve essere scritto nel suo programma senza ambiguità; tuttavia, l'interpretazione delle regole dipende dall'individuo. I dati che il programma può ottenere sono un gruppo ben delimitato scelto per la sua rilevanza rispetto agli scopi del programma stesso; ma è il programmatore, naturalmente, che decide in merito e che sceglie i dati. Così, nelle situazioni in cui la risposta automatica è tutto ciò che viene richiesto, ovvero dove gli oggetti sono ben definiti, l'insieme di dati 'rilevanti' è circoscritto

149 Cfr. Paternoster 2002 pp. 207-211.
150 Cfr. Churchland 1989 tr.. it. pp. 14-15.
151 Cfr. Dreyfus 1972.

chiaramente: si può procedere usando programmi di risposta derivati da esperienze passate. È in questi campi di attività, dove non si richiede pensiero creativo e complesso, che la logica formale e il computer possono comportarsi abbastanza bene. Ma tali casi sono piuttosto rari; infatti esistono capacità in ogni manifestazione del comportamento intelligente umano che non sono programmabili:

> il comportamento dell'uomo non può essere spiegabile in termini di un meccanismo di trattamento delle informazioni che elabora impulsi che rappresentano caratteristiche del mondo. Nulla che provenga dalla fisica o dall'esperienza suggerisce che le azioni dell'uomo possano essere così spiegate[152].

Paternoster afferma che comunque il funzionalismo computazionale non ha completamente fallito poiché, partendo proprio dalle critiche che gli sono state mosse, ha dato vita ad altre ricerche che, seppur negando i principi della TCRM, sembrano fortemente legate con il funzionalismo computazionale. Si tratta, per esempio, del connessionismo e della teoria dei sistemi dinamici.

Il *connessionismo* è un programma di ricerca che si fonda sull'idea che i processi cognitivi sono particolari tipi di computazioni eseguite da reti neurali, ossia da strutture formate da unità interconnesse, ciascuna delle quali simula, in maniera astratta, il comportamento di un neurone o un complesso di neuroni.

La *teoria dei sistemi dinamici* invece si basa sul fatto che i modelli matematici non lineari descrivono l'evoluzione di sistemi complessi[153]. Si ha una descrizione matematica capace di generare previsioni sul comportamento date certe condizioni iniziali.

Nonostante le critiche che gli sono state mosse, e nonostante anche il suo fondatore, Putnam, alla fine abbia preso le distanze da esso, non si può negare l'importanza del funzionalismo, il quale è alla base dell'odierna intelligenza artificiale, fondata anch'essa sul presupposto che se gli stessi processi possono essere realizzati da vari supporti materiali, niente esclude che processi come quelli eseguiti dalla mente umana possano essere realizzati anche da una macchina.

---

152 Ivi, p. 267.

153 Cfr. Paternoster 2002, p. 212.

1.6. Il punto focale del dibattito: la sopravvenienza

In questo capitolo si sono illustrate alcune soluzioni al problema mente-corpo. In quel che segue ci focalizzeremo sulla soluzione che sembra offrirci l'argomento della sopravvenienza, introdotto nell'ambito del dibattito mente-corpo da Davidson e che tanto ha animato il discorso filosofico degli ultimi decenni.

La relazione di sopravvenienza, come ci fa notare Stalnaker, «is a relation between sets of properties or facts, […] a supervenience claim says only that all of the properties from one set somehow determine the properties of the other»[154]; o usando le parole di David Lewis «Supervenience means that there could be no difference of one sort without difference of the other sort»[155].

Nel pensiero di Davidson, si è visto, l'insieme di proprietà o fatti di cui si parla sono quelli mentali e fisici, la sopravvenienza in Davidson dunque rende conto della relazione tra mentale e fisico senza voler ridurre le proprietà mentali a mere proprietà fisiche e senza volerle eliminare completamente. L'idea della sopravvenienza, dunque, ci dovrebbe aiutare a fornire una versione non riduzionista di materialismo. Kim, filosofo che tanto si è occupato di tale argomento, ritiene che «supervenience is to be consistent with the irreducibility of supervenient properties to their base properties»[156].

Tuttavia, sembra che la sopravvenienza sollevi più problemi di quanti ne risolva: basti pensare che, accanto alla versione "debole" di sopravvenienza avanzata da Davidson, vedremo anche l'ipotesi di sopravvenienza "forte" (o causazione sopravveniente) per quanto concerne Kim, di sopravvenienza "globale" in Putnam e addirittura di "a-sopravvenienza" in White. Giusto per citare un altro tipo di visione della sopravvenienza, ma di cui non tratteremo nella ricerca, Terence Horgan parla di *superdupervenience*, intendendo con questo termine una «ontological supervenience that is robustly explainable in a materialistically acceptable way»[157], ossia una sopravvenienza ontologica completamente spiegabile in un

154 Stalnaker 1996, p. 224.
155 Lewis 1986, p. 15.
156 Kim 1993b, p. 140.
157 Horgan 1993, p. 566.

modo materialistico[158]. Ciò per sottolineare le evidenti difficoltà di trovare una descrizione della nozione di sopravvenienza che sembri soddisfacente.

Quel che mostreremo nei capitoli successivi sono quindi le problematiche che Kim, Putnam e White trovano nell'argomento della sopravvenienza, in relazione a come formulare una tesi convincente della sopravvenienza, a quale efficacia esplicativa e causale hanno le proprietà mentali e a quale relazione tra proprietà mentali e fisiche si viene così a creare, presentando anche le soluzioni proposte da questi nell'avanzare la loro versione di tale nozione.

CAPITOLO 2

Jaegwon Kim: la sopravvenienza e il ruolo delle proprietà mentali

[158] Secondo Horgan il problema della sopravvenienza risiede nella spiegazione delle relazioni a cui tale nozione dà luogo, relazioni che sembrano essere fondamentali e inspiegabili, portando all'utilizzazione di categorie esplicative a loro volta non spiegabili.

2.1. Le conseguenze della visione davidsoniana: epifenomenismo e eliminativismo

Il monismo anomalo di Davidson ha subito numerose critiche, tra le quali le più incisive sono state quelle di *epifenomenismo* ed *eliminativismo*; come accennato in precedenza prenderemo in esame fondamentalmente le critiche rivoltegli da Kim.

La prima critica si incentra sul fatto che l'evento mentale così come concepito dal monismo anomalo pare essere causa non in quanto mentale, bensì in quanto fisico, e questo nega l'efficacia causale degli stati mentali, cadendo appunto nell'epifenomenismo. Con questo termine ci si riferisce all'apparente irrilevanza causale degli stati mentali: il mentale non ha alcun effetto sul fisico e quindi è causalmente inerte (epifenomeno), poiché gli eventi mentali, che comunque possono essere causati da fenomeni fisici, non possono causare nulla, né eventi mentali né eventi fisici. Nel monismo anomalo, si ricorderà, gli eventi mentali sono eventi fisici, e in base al carattere nomologico della causalità gli eventi sono cause ed effetti solo se istanziano leggi fisiche; tuttavia, e in questo risiede l'anomalia del mentale, il mentale non è governato da leggi strette; in più, non ci sono leggi che connettono il mentale al fisico.

Kim ritiene che in Davidson l'identità tra eventi mentali ed eventi fisici risulta essere un fatto inspiegabile: da un lato c'è il mondo della fisica e delle relazioni causali; dall'altro il mentale, il quale non sarebbe che il mondo della fisica ma osservato da un'altra prospettiva. Se le cose stanno effettivamente così, continua Kim, non si capisce come si possa avere identità tra eventi mentali ed eventi fisici, anche perché Davidson sostiene che non ci sono

leggi che correlano i due ambiti. Dunque Kim conclude che il monismo anomalo non ci dice nulla sul problema mente-corpo, cioè su come è articolata la relazione tra mentale e fisico. Ritenere che non ci sono leggi che collegano i generi mentali, o le proprietà mentali, con quelli fisici non ci dice come i due generi o le due proprietà sono relazionate, quindi non ci dice niente sul rapporto mente-corpo[159]. Egli perciò afferma che «we want our mind-body theories to tell us more […] about how mental properties and physical properties *are* related, and hopefully also explain why they are so related. We don't get such a story from anomalous monism»[160].

In realtà Davidson ritiene di aver spiegato tale rapporto attraverso il concetto di sopravvenienza, ed è proprio su tale concezione che si centra la critica di Kim. L'identità tra mentale e fisico, così come intesa da Davidson, non sarebbe sufficiente secondo Kim per dare agli stati mentali quell'efficacia causale che invece egli vuole attribuirgli.

2.1.1. L'esclusione causale

Un problema derivante da quello della sopravvenienza (e problema principale della causazione mentale[161]) è quello dell'*esclusione causale.* Davidson viene accusato da Kim del fatto che la sua concezione degli stati mentali violerebbe un principio che ogni fisicalista dovrebbe accettare, ovvero appunto il cosiddetto principio dell'esclusione causale, secondo il quale data una relazione causale, essa non può essere determinata da due eventi indipendenti e singolarmente sufficienti – e dunque uno non può che escludere l'altro – a meno di effettiva sovradeterminazione[162].

La sovradeterminazione ammette l'eventualità che si dia più di un tipo di causa rispetto a un effetto, ovvero ammette l'eventualità che più cause indipendenti tra loro e sufficienti per lo stesso effetto abbiano luogo nel medesimo istante: l'effetto risulterebbe così sovradeterminato, ossia determinato da più di una causa. Tuttavia la sovradeterminazione in genere riguarda lo stesso tipo di cause, per esempio due cause fisiche. Supponiamo che due cecchini sparino contemporaneamente la stessa persona in due punti vitali diversi. La morte

159 Cfr. Kim 1998, p. 4.
160 Ivi, p. 5.
161 Kim individua tre problemi fondamentali insiti nella tesi della causazione mentale: 1. l'anomalia del mentale; 2. le proprietà mentali estrinseche; 3. l'esclusione causale. Tuttavia solo l'ultimo dei tre è strettamente legato con qualsiasi posizione materialista e quindi anche con il monismo anomalo.
162 Cfr. Kim 2005 p. 42, cfr. anche Gozzano 2007, p. 100.

della persona è sovradeterminata, ovvero determinata da due cause (i due diversi colpi sparati) che però in questo caso sono dello stesso genere. Ciò che però in questo caso il principio di esclusione causale vuole escludere è piuttosto la sovradeterminazione tra tipi diversi di cause, quindi ciò che ci preme sapere è se dato un effetto, poniamo sollevare un braccio, è l'evento mentale (e quindi la decisione di alzare il braccio) o l'evento fisico (il fatto di trovarsi in un determinato stato cerebrale) a causare l'effetto di sollevare il braccio.

Accettare la causazione mentale come fa Davidson, argomenta Kim, significa accettare che un evento mentale *m* (per esempio avere mal di denti) possa causare un evento fisico *p* (per esempio prendere un'aspirina), ma a meno che non si voglia ricadere in una sorta di dualismo, bisogna ammettere che esiste anche una causa fisica *p** per *p*; in base alla sopravvenienza del mentale sul fisico infatti ad ogni causa mentale è possibile associare una causa fisica. La presenza di due cause (*m* e *p** in questo caso), ognuna delle quali offre un resoconto causale completo di un dato evento (*p*), crea una situazione instabile che ci richiede di trovare un modo per spiegare come le due presunte cause sono relazionate l'una con l'altra. Questo è il problema dell'esclusione causale dice Kim, non si capisce come sia possibile per la mente esercitare i suoi poteri causali in un mondo che è essenzialmente fisico:

> the problem of causal exclusion is to answer this question: *Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?*[163].

Davidson ha replicato a questa critica affermando che essa confonderebbe le *relazioni causali*, le quali sussistono tra gli eventi in modo indipendente da come vengono individuati, con le *spiegazioni causali*, che invece correlano descrizioni. Nelle spiegazioni causali gli eventi vi entrano in quanto mentali, ovvero in una descrizione che fa uso di un vocabolario mentale. Nelle relazioni causali invece gli eventi entrano in quanto tali, poiché un evento è ciò che è e le sue proprietà causali non dipendono da come lo descriviamo. Quindi in questo senso, un evento individuato mentalmente può essere una causa. Lo può essere in questo modo: è un singolo evento particolare, un'occorrenza appunto, che causa un altro singolo evento particolare (un'altra occorrenza).

[163] Kim 2005, p. 38.

Tuttavia Kim sottolinea come Davidson riconosca alle proprietà rilevanza causale: la tesi della sopravvenienza, così come intesa da Davidson, ci dice infatti che se due eventi hanno proprietà mentali diverse sono diverse anche le loro proprietà fisiche, che sono quelle causalmente efficaci. Ma questo tipo di concezione non fornisce efficacia causale alle proprietà mentali.

Riportiamo un esempio fornito da De Caro che ci pare esplicativo al riguardo: immaginiamo che ci siano due tipi di automobili chiamate rispettivamente *Velox* e *Tardus*. Le *Velox* hanno una velocità massima che varia dai 200 ai 250 km/h in base al modello, e sono disponibili solo nel colore rosso. Le *Tardus* invece hanno una velocità massima che varia dai 100 ai 150 km/h e l'unico colore disponibile è il grigio. Si può affermare che il colore delle auto sopravviene sulla proprietà della loro velocità, poiché se due macchine sono di colore diverso sarà diversa anche la loro velocità (un'auto rossa e una grigia non raggiungeranno mai la stessa velocità massima). Tuttavia non è detto il contrario, cioè due macchine con velocità massime diverse possono essere dello stesso colore. Nel caso di queste automobili, dunque, il colore sopravviene sulla velocità. Ma quale sarà la proprietà che, ipotizziamo, farebbe vincere una gara automobilistica a una di queste macchine: il colore o la velocità? Ovviamente le auto rosse vinceranno sempre sulle auto grigie, dunque possiamo affermare che il colore è causalmente rilevante; ma la causa della vittoria (o sconfitta) sarà chiaramente la velocità, quindi in questo caso il colore non è causalmente efficace. La nozione centrale quando si parla di eventi in rapporto causale reciproco è quella di rilevanza e non quella di efficacia: la *Velox* vince a causa del fatto che è più veloce, non perché è rossa[164].

Le proprietà mentali nel monismo anomalo equivarrebbero ai colori nell'esempio citato: esse sono causalmente efficaci, ma non rilevanti.

L'esclusione causale dimostra, secondo Kim, l'inevitabilità di una qualche forma di riduzione del mentale al fisico. La sopravvenienza darebbe solo l'illusione di poter fare a meno della riduzione. Parlare di sopravvenienza del mentale sul fisico significa ridurre il potere causale del mentale: questo è il problema dell'esclusione causale, che rischia di relegare il mentale allo statuto di epifenomeno.

Secondo Kim quindi non è possibile conciliare la tesi dell'efficacia causale degli stati mentali con la sopravvenienza o altre tesi antiriduzionistiche equivalenti, se non cadendo nell'epifenomenismo o nella sovradeterminazione causale.

[164] Cfr. De Caro 1998, pp. 121-2.

Oltre al principio di esclusione causale Kim, nella sua critica al monismo anomalo, prende in considerazione anche un altro principio, secondo lui ancor più decisivo nel mostrare la fallacia del pensiero di Davidson, ovvero il principio della chiusura causale.

2.1.2. La chiusura causale

Stando alla tesi della chiusura causale, ogni evento fisico che ha una causa al tempo *t* ha una causa fisica a *t*[165]. Kim ritiene che la relazione di identità di occorrenze e la relazione di sopravvenienza siano troppo deboli per ottenere ciò che Davidson vorrebbe, ovvero difendere la chiusura causale del mondo fisico e al contempo l'efficacia causale degli stati mentali. Al fine di ottenere ciò per Kim sono necessarie delle leggi psico-fisiche, leggi che invece Davidson, come abbiamo visto, sostiene non esistono:

> if mental properties are physically irreducible and remain outside the physical domain, then, given that the physical domain is causally closed, how can they exercise causal powers, or enjoy any kind of causal relevance, in the physical domain?[166]

Dato questo principio, argomenta Kim, non si capisce come le proprietà mentali possano avere un ruolo, a meno di non ritornare al dualismo cartesiano, ovvero alla concezione che qualche evento fisico ha bisogno di cause non fisiche. E questo significa per Kim che il nostro mondo fisico si deve riferire a cause non fisiche, cosa che la maggior parte dei fisicalisti troverebbe inaccettabile[167].

Kim definisce il principio di chiusura causale così: «If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical»[168]. Ciò che sostiene è quindi che non ci possono essere influenze causali non fisiche introdotte dentro il dominio fisico dall'esterno. La chiusura causale quindi, continua Kim, non ci dice che gli eventi e le entità fisiche sono tutto ciò che c'è in questo mondo, o che

165 Cfr. Kim 1993, p. 280.
166 Kim 1998, pp. 58-9.
167 Cfr. Kim 1993, p. 280.
168 Kim 1998, p. 40. Cfr. anche Kim 2005, pp. 15-6; Kim 1993, p. 280.

la causazione si esaurisce nella causazione fisica. Ci possono essere entità ed eventi al di fuori del mondo fisico, e tra di loro possono anche intercorrere delle relazioni causali, tuttavia non ci possono essere influenze causali che arrivano dentro al dominio fisico dall'esterno.

Ogni evento mentale dunque deve avere una base fisica, e se ogni proprietà mentale viene istanziata perché, e solo perché, uno dei suoi realizzatori fisici è istanziato, allora c'è una dipendenza analoga delle occorrenze mentali dalle occorrenze fisiche. La sopravvenienza porta dunque i fenomeni mentali all'interno dell'ambito fisico: il fisico determina il mentale, e in questo senso il mentale non costituisce un dominio ontologico indipendente[169].

Seguiamo il ragionamento di Kim: supponiamo che un certo evento, in virtù delle sue proprietà mentali, causi un evento fisico. La chiusura causale del mondo fisico ci dice che questo evento deve avere anche una causa fisica. Assumiamo quindi che questa causa fisica, in virtù delle sue proprietà fisiche, causa l'evento fisico. Ora ci si pone la domanda di quale sia la relazione tra la causa fisica e quella mentale, poiché entrambe sembrano cause dell'effetto fisico. Kim ritiene che si possano prendere due strade, la prima sarebbe supporre che ciascuna delle cause è solo una causa parziale, ma che assieme diventano una causa sufficiente: «mental cause and physical cause are each only a partial cause, and that they together make up one sufficient cause»[170]. Questa strada però non pare quella giusta poiché viola il principio della chiusura causale in quanto l'evento mentale è un componente necessario per una causa completa. Un'altra possibilità, continua Kim, è quella di vedere le due cause come indipendenti per l'effetto fisico. Ma in questo caso sembra che ci si trovi davanti a un caso di sovradeterminazione. Infatti, se l'evento fisico non fosse successo, la causa mentale da sola avrebbe causato l'effetto. Ma questo non è possibile se teniamo conto, come un fisicalista deve fare, del principio della chiusura causale, poiché se supponiamo che l'evento fisico non sia successo stiamo violando tale principio.

Kim ritiene che sia proprio il principio della chiusura causale a escludere la causa mentale facendo prevalere quella fisica, mentre l'esclusione causale sarebbe neutrale rispetto alla competizione fisico-mentale. L'esclusione causale infatti sostiene che deve esserci o la causa mentale o quella fisica, ma non favorisce né l'una né l'altra. Anche il principio della chiusura causale in linea di principio non esclude che ci siano cause non fisiche. Infatti, argomenta Kim, se la situazione della chiusura causale fosse al contrario, ovvero fosse il

169 Cfr. Kim 1998, p. 41.
170 Kim 1993, p. 280.

dominio mentale a essere causalmente chiuso e non quello fisico, la causa mentale avrebbe prevalso su quella fisica. Sarebbe però l'unione dei due principi che secondo Kim porta all'epifenomenismo.

Facciamo un esempio per chiarire meglio questo punto: supponiamo che l'istanziazione di una proprietà mentale M causi l'istanziazione di un altro evento mentale M*: per esempio M è avere mal di testa e M* l'intenzione di prendere un'aspirina. In base alla sopravvenienza M* ha una causa fisica P*. Ma qual è la causa per M*? M o P*? In base alla sopravvenienza M* occorre perché occorre la sua base sopravveniente P*, e finché P* occorre anche M* occorre, a prescindere dagli altri eventi che precedono l'istanziazione di M*, in particolare a prescindere all'avvento o meno di M. Anche M tra l'altro, sempre in base al concetto di sopravvenienza, ha una base fisica P. Cosa causa P*, M o P? Affinché M causi M* è necessario che M causi P* portando alla sovradeterminazione causale: pare infatti che P* sia causato sia da P che da M. Se P è sufficiente perché M occorra, e M è sufficiente per P* allora, conclude Kim, P è sufficiente per P*. E qui entra in gioco il principio della chiusura causale (che, lo ricordiamo, sostiene che un evento fisico non può avere cause che non siano fisiche, escludendo così M), che per evitare la sovradeterminazione ammette la priorità di P rispetto a M e quindi la base mentale perde qualsiasi ruolo rilevante diventando un epifenomeno[171].

Come sono dunque relazionate tra loro le cause mentali e quelle fisiche, com'è possibile la causazione mentale se ogni evento fisico ha una causa fisica? Questo è quello che Kim chiama "the problem of causal-explanatory exclusion":

> The problem seems to arise from the fact that a cause, or causal explanation, of an event, when it is regarded as a full, sufficient cause or explanation, appears to *exclude* other *independent* purported causes or causal explanations of it[172].

Non ci occuperemo per ora della soluzione proposta da Kim, ma ci limitiamo a esporre la sua critica nei confronti di Davidson.

Quest'ultimo replica che non ha senso chiedersi se un evento ha un particolare effetto in virtù del suo essere mentale o del suo essere fisico, sarebbe come chiedersi se il suo effetto

[171] Cfr. Kim 1998, pp. 41-6.
[172] Kim 1993, p. 281.

deriva dal fatto di essere descritto in italiano o in inglese. La fisica è l'attività di scoprire la profonda struttura causale del mondo, tuttavia essa non ha il monopolio sul modo di descrivere ciò che succede a questo mondo.

2.1.3. Determinazione e covarianza

Davidson, sostiene Kim, è solito associare l'idea di sopravvenienza con quella di *dipendenza* o *determinazione*: se il mentale sopravviene sul fisico esso è dipendente dal fisico, o il fisico determina il mentale, nel senso che la natura mentale di una cosa è interamente fissata dalla sua natura fisica.

La tesi che un dato dominio sopravviene su un altro è una tesi metafisica su un'oggettiva relazione di dipendenza tra i due domini, ma non ci dice niente riguardo se o come i particolari della relazione di dipendenza possono essere conosciuti così da permettere la formulazione di spiegazioni, riduzioni o definizioni.

La sopravvenienza debole sostenuta da Davidson, afferma Kim, è troppo debole per molte delle sue applicazioni e priva il mentale di ogni potere causale. Quello che mette in luce è solo una relazione parziale di determinazione. Il primo motivo per il quale Davidson usa la sopravvenienza psicofisica è quello di riconoscere la dipendenza del mentale dal fisico ma allo stesso tempo nega che ci siano leggi che connettono le proprietà psicologiche e fisiche. Ma se si vuole la dipendenza psicologica, avverte Kim, ci si deve preparare per le leggi psicofisiche.

Quando abbiamo parlato della sopravvenienza in Davidson abbiamo detto che questa viene definita "debole" e che presenta tre componenti: covarianza, dipendenza e irriducibilità. Kim ritiene che la covarianza sia il "componente critico" poiché qualsiasi concetto di sopravvenienza (sia che si riferisca al rapporto mente-corpo che alla morale) deve includerlo: il problema della sopravvenienza riguarda quindi la relazione tra la covarianza e gli altri due componenti. Ciò che a questo punto Kim si chiede è se «is there an interpretation of covariance that is strong enough to sustain supervenience as a dependency relation but weak enough not to imply reducibility»[173].

[173] Kim 1993, p. 140.

La nozione di covarianza proposta da Davidson, secondo Kim, pretende di generare tra le proprietà mentali e quelle fisiche una relazione di dipendenza. Egli definisce questo tipo di covarianza "debole":

> *Weak covariance:* No possible world contains things, x and y, such that x and y are indiscernible in respect of properties in B ("B-indiscernible") and yet discernible in respect of properties in A ("A-discernible")[174].

Secondo la nozione di covarianza utilizzata da Davidson quindi all'interno dello stesso mondo, se due elementi (in questo caso x e y) sono indiscernibili per quanto riguarda le B proprietà, essi lo saranno anche per le A proprietà. O, detto diversamente, se due individui sono indiscernibili per quanto riguarda tutte le proprietà mentali lo saranno anche riguardo le proprietà fisiche, sempre fermo restando che questa relazione non vale in tutti i mondi possibili.

Kim ritiene questa una relazione inadatta a qualsiasi tesi di dipendenza con forza modale, poiché la condizione vincolante opera solo entro un mondo singolo per volta. Il fatto che il mentale sia distribuito in un certo modo in un mondo non ha nessun effetto sul come potrebbe essere distribuito in altri mondi possibili.

La causalità mentale, afferma Kim, non è per niente garantita dalla sopravvenienza debole, infatti la rete di relazioni causali che interconnettono gli eventi è ciò che dà una struttura intellegibile a questo universo di eventi. Tuttavia, continua Kim, nella formulazione di Davidson, ciò che è garantito è la corrispondenza coerente per ogni mondo possibile tra il fisico e il mentale, non l'uguaglianza di questa relazione in tutti i mondi. La sopravvenienza debole (e con essa la covarianza debole) dipende dal particolare mondo preso in considerazione, e non è una caratteristica invariante attraverso mondi possibili.

Per esempio, spiega Kim, in un mondo fisicamente uguale al nostro potrebbe non esservi alcun carattere morale: se spostiamo un solo neutrone del mondo *a*, avremo un mondo *b* dai caratteri morali totalmente imprevedibili. Nel nostro mondo (supponiamo il mondo *a*), per esempio, essere coraggioso, benevolo e onesto (poniamo A-proprietà), sono considerati tratti di una persona che potremo definire buona(poniamo B-proprietà). Ma nel mondo *b* tutte queste caratteristiche invece potrebbero corrispondere a quelle di una persona cattiva. Tutta la

[174] Ivi, p. 141.

morale, mondo per mondo, si ridurrebbe non a ciò che è giusto, ma a ciò che è coerente, e questo perché in un mondo possibile una persona o un oggetto che ha le stesse A-proprietà che ha in questo mondo (per esempio essere coraggioso, onesto,ecc) non deve necessariamente avere le stesse B-proprietà (ovvero essere buono) che avrebbe nel nostro mondo. Il suo principio di universalizzazione dell'azione diventerebbe una mera esigenza di coerenza generalizzata. Se essere un uomo buono dipende, o è determinato, da certi tratti di carattere, allora avere questi tratti deve assicurare o garantire il fatto di essere un uomo buono[175].

Lo stesso discorso vale per la sopravvenienza psico-fisica in connessione alla questione della causalità mentale. Per questo Kim afferma che «*the very same network of causal relations would obtain in Davidson's world if you were to redistribute mental properties over its events any way you like; you would not disturb a single causal relation if you randomly and arbitrarily reassigned mental properties to events, or even removed mentality from the world.* The fact is that under Davidson's anomalous monism, mentality does no causal work»[176].

Il fatto che il mentale possa, mondo per mondo, essere ridistribuito a piacere sul fisico, o addirittura eliminato del tutto, dimostra che esso è privo di qualunque ruolo causale, poiché questo tipo di covarianza non genera dipendenza.

La posizione di Davidson, conclude Kim, finisce così per assimilarsi a quell'eliminativismo che proprio l'autore pretendeva rifiutare:

> I believe that if we push anomalous monism this way, we will find that it is a doctrine virtually indistinguishable from outright eliminativism. […] Anomalous monism, rather than giving us a form of nonreductive physicalism, is essentially a form of eliminativism.[177]

Dunque, ciò che Davidson considera come autonomia psicologica della mente, per Kim altro non è che una forma di eliminativismo. L'idea della sopravvenienza degli stati mentali proposta da Davidson non può sostenersi senza cadere nel riduzionismo o nel dualismo.

---

175 Ivi, p. 60.
176 Ivi, p. 269, cfr. anche Kim 1998, p. 34.
177 Kim 1993b, p. 270.

2.2. La soluzione di Kim

Per capire la posizione di Kim dobbiamo innanzitutto prendere in considerazione la sua critica al materialismo non riduzionista e all'argomento della realizzabilità multipla.

Come illustrato nel capitolo precedente, Kim ritiene di aver dimostrato tramite il principio di chiusura causale e quello dell'esclusione causale che sia il dualismo (nella versione delle proprietà e delle sostanze), sia una tesi antiriduzionista come il monismo anomalo, non possono funzionare. Infatti, egli sostiene che l'antiriduzionismo nega alla mente poteri causali e la relega a un'entità isolata, la cui presenza (o assenza) non farebbe alcuna differenza causale: nel voler salvare il mentale come qualcosa di distinto e speciale finisce invece per privarlo di qualsiasi potere causale.

Kim ritiene che il fisicalismo non riduzionista abbia sempre riscosso un certo fascino perché salvaguarda sia l'autonomia delle scienze speciali, come la psicologia, sia il nostro essere agenti cognitivi. Tuttavia, sottolinea Kim, esso non ci può spiegare come i fenomeni mentali possono influire causalmente sul corso degli eventi nel mondo fisico. E siccome non è in grado di spiegare la causazione mentale, egli ritiene che sia una tesi incompleta e inadeguata:

> Mental causation is fundamental to our conception of mentality and to our view of ourselves as agents and cognizers; any theory of mind that is not able to accommodate mental causation must be considered inadequate, or at least incomplete[178].

La soluzione per salvare la causalità mentale per Kim risiede invece proprio nel riduzionismo; questo infatti può mantenere l'efficacia del mentale al costo della sua autonomia: il mentale può trovare posto all'interno del dominio fisico, «our only hope […] if mental causation is to be saved, is physical reductionism. […] If mental is to have a causal influence in the physical domain—in fact, if it is to have any causal efficacy at all—it must be physically reducible»[179].

[178] Kim 2005, p. 153.
[179] Ivi, pp. 159-161.

Bisogna tuttavia stare attenti a non confondere il riduzionismo con l'eliminativismo, avverte Kim: «witches were eliminated, not reduced; the gene has been reduced, not eliminated»[180]. Egli ritiene che dovremo verificare se cose come credenze e desideri sono riducibili a processi e proprietà neurali, biologici o meno.

2.2.1. Le proprietà mentali

Kim sostiene la tesi secondo cui la causazione mentale necessita la riduzione, e che chiunque crede nella causazione mentale deve essere preparato ad abbracciare il riduzionismo mente-corpo. La riduzione quindi è per lui fondamentale per l'efficacia causale del mentale.

Tuttavia il mentale, prosegue Kim, non è una totalità che può essere o tutta irriducibile o tutta riducibile; egli infatti ritiene che non tutto il mentale può essere reso nei termini della fisica. La parte del mentale che resta fuori da questa riduzione sono i cosiddetti *qualia*: solo questi sarebbero per Kim irriducibili. Egli porta gli esempi di Chalmers e Nagel come pensatori che come lui ritengono che il carattere qualitativo di un'esperienza cosciente non può essere ridotto. Spiegheremo meglio più avanti questo punto.

Nella sua critica all'antiriduzionismo Kim prende in considerazione la *realizzabilità multipla*, argomento influente tra i filosofi non riduzionisti a sostegno di una visione autonoma del mentale, argomento sostenuto in primis da Putnam[181]. Questi, lo ricordiamo, sosteneva contro il riduzionismo (in particolare quello della teoria dell'identità di tipo) che uno stesso stato mentale può essere realizzato da diversi stati fisici, garantendo un'autonomia funzionale ed esplicativa agli stati e processi mentali nei termini dei loro ruoli causali, rispetto alle loro basi potenzialmente diverse di realizzazione fisica.

Quindi non esisterebbe un singolo stato neurale, poniamo N, che realizza il dolore in tutti i tipi di organismi o di sistemi fisici, piuttosto c'è una molteplicità di generi di stati fisici, es. N_1, N_2, N_3, ecc., tali che N_1 realizza il dolore negli esseri umani, N_2 nei rettili, N_3 in ipotetici marziani e così via. Il punto centrale della realizzabilità multipla contro il riduzionismo Kim lo individua nel fatto che le proprietà mentali non hanno proprietà fisiche nomicamente coestensive, tuttavia egli ritiene che la base di realizzazione fisica per gli stati mentali, così come caratterizzata dalla realizzabilità multipla, debba essere relativizzata alla

[180] Ivi, p. 160.
[181] Cfr. Putnam 1975, cap. 21, *The Nature of Mental States*.

specie:

> The multiple realization argument perhaps shows that the strong connectibility of mental properties vis-à-vis physical properties does not obtain: however, it *presupposes* that *species-specific strong connectibility* does hold. [...] The phenomenon of multiple realization is consistent with the species-specific strong connectibility[182].

Ciò che Kim sostanzialmente afferma è che anche se non tutti i tipi, per esempio di dolore, sono riducibili a un unico tipo di stato fisico globale, che valga in assoluto per tutte le specie, il dolore degli esseri umani, o quello dei rettili e così via, è comunque riducibile a un unico stato fisico in quella specie, quindi abbiamo comunque una riduzione specie-specifica di uno stato mentale a uno stato fisico anche nella realizzabilità multipla.

La proposta di Kim invece è quella di vedere le proprietà mentali come *disgiuntive*, ovvero proprietà che consistono di o una proprietà fisica, o una seconda proprietà fisica, e così via; possiamo dare una definizione di dolore, per esempio, come essere l'hardware umano del dolore, o l'hardware dei robot per il dolore, ecc., dunque considerare la disgiunzione come singolo sostrato fisico di uno stato mentale.

Per avvalorare la sua posizione Kim riporta l'esempio della giada[183]: questa, contrariamente a quanto si pensava un tempo, non è un unico minerale, ma è composto da due distinti minerali, giadeite e nefrite. Ora prendiamo in considerazione l'affermazione (L) "La giada è verde". Prima della scoperta della sua duplice natura quest'affermazione era una legge, ora però sappiamo che (L) è la congiunzione di due leggi: (L_1) "La giadeite è verde"; (L_2) "La nefrite è verde". Cosa succede con (L)? Sembrerebbe che possa continuare ad essere una legge poiché supportata dai controfattuali: se qualcosa è giada, se qualcosa è un campione di giadeite o nefrite, allora in entrambi i casi, segue che è verde. Ma, prosegue Kim, c'è un aspetto che questo esempio non supporta, e cioè la "proiettabilità", ovvero l'abilità di essere confermato da esempi positivi. Supponiamo di riesaminare i campioni di giada che possediamo dopo aver scoperto la sua duplice natura, e scopriamo che quelli che possediamo sono tutti campioni di giadeite e nessuno di nefrite; se questo dovesse succedere non possiamo continuare a ritenere che (L) sia confermata. Ciò che è stato confermato è solo (L_1),

182 Kim 1989, p. 38.
183 Kim 1993b, pp. 319-22; cfr. anche Kim 2006, pp. 185-87.

che non ha niente a che fare con (L_2); (L) è la congiunzione di due leggi, una confermata e l'altra no, tuttavia tutti i campioni verdi di giadeite sono esempi positivi di (L). Il fatto che (L) non sia confermata come in genere sono confermate le leggi è dovuto al fatto che la giada è un genere disgiunto, una disgiunzione di due generi nomici eterogenei, la quale però non è essa stessa un genere nomico. Dunque, conclude Kim, pensiamo alla giadeite e alla nefrite come generi distinti perché sono generi chimici differenti; la distinzione chimica è importante perché le proprietà dei minerali sono sopravvenienti sulla loro microstruttura, le proprietà macrofisiche sono determinate dalla microstruttura.

Tornando al nostro esempio del dolore, esso occorre quando alcune condizioni fisiche sono presenti nel sistema; le relazioni nomiche tra le proprietà degli stati mentali sono dovute e spiegabili nei termini delle proprietà e delle connessioni nomiche-causali tra i loro substrati fisici. Pare che la situazione sia parallela a quella della giada, avverte Kim. Sembra che pensiamo alla giadeite e alla nefrite come generi diversi, e alla giada non come a un genere. E questo succede perché giadeite e nefrite sono generi chimici diversi. Egli quindi si chiede perché la relazione tra il dolore e le sue basi realizzatrici (N_1, N_2, N_3, ecc.) non è uguale a quella tra giada e la relazione tra giadeite e nefrite? La proprietà di avere una delle proprietà N_1, N_2, N_3 e così via, è identica alla proprietà disgiuntiva N_1 v N_2 v N_3 v ecc. Le proprietà mentali sono disgiunzioni delle loro basi di realizzazione fisica.

Nello stesso modo in cui non abbiamo teorie scientifiche per la giada (se proprio ne volessimo una, dice Kim, potremmo aiutarci con le congiunzioni della teoria della giadeite e della teoria della nefrite), non avremo una teoria psicologica unitaria, bensì una teoria per quanto riguarda il dolore negli esseri umani, la teoria per il dolore nei rettili e così via. Quindi avremo solo una congiunzione di teorie per ogni specie biologica e per ogni struttura fisica.

Questo diventa chiaro, argomenta Kim, se teniamo in considerazione due principi: il *principio dell'individuazione causale dei generi* e il *principio di eredità causale.* Il primo principio sostiene che:

> Kinds in science are individuated on the basis of causal powers; that is, objects and events fall under a kind, o share in a property, insofar as they have similar causal powers[184].

[184] Ivi, p. 190.

Secondo Kim questo ci fa capire che le diverse istanziazioni del dolore (e quindi N_1, N_2, N_3, ecc.) sono eterogenee sì riguardo al genere, ma esse sono anche diverse riguardo ai poteri causali e quindi entrano in diverse leggi causali.

In base al principio di eredità causale invece, se la proprietà mentale M è realizzata in un sistema al tempo *t* in virtù della base fisica di realizzazione P, i poteri causali di questa istanziazione di M sono identici ai poteri causali di P:

> the causal powers of an instance of a second-order property are identical with (or a subset of) the causal powers of the first-order realizer that is instantiated on that occasion[185].

I funzionalisti per bloccare questo tipo di caratterizzazione hanno affermato che le proprietà disgiuntive non sono autentiche proprietà scientifiche poiché, per esempio, non compaiono in leggi fisiche. Kim tuttavia sostiene che, a suo avviso, anche gli stati funzionali sono non scientifici: se uno stato mentale come il dolore è realizzato in differenti strutture fisiche, allora essi possono essere scartati come oggettivi oggetti scientifici; pertanto, se la realizzabilità multipla è vera i generi mentali non sono scientifici.

Riassumendo, Kim ritiene che ogni esemplificazione di un certo genere mentale (il dolore) si identifica alternativamente, disgiuntivamente, con un'esemplificazione o di P_1 o di P_2, ecc., dove P_1 e P_2, indicano i singoli realizzatori fisico-neurali specie-specifici del genere mentale M, ossia relativi alle singole specie prese in considerazione[186]. Questo perché i realizzatori fisico-neurali del genere mentale M (dolore) cambiano al variare delle specie prese in considerazione, così che il dolore avrà diversi realizzatori fisico-neurali che si tratti del dolore nella specie degli esseri umani (P_1), dei rettili (P_2), e così via. Quindi la singola esemplificazione del genere mentale M si identifica alternativamente, cioè in modo specie-specifico, con una delle singole esemplificazioni del genere P, ed è proprio la base fisico-neurale specie-specifica a rendere il dolore umano, ad esempio, diverso dal dolore rettiliano, ecc.

Le proprietà mentali sono quindi solo localmente[187] (e non globalmente) nomiche, ossia conformi a leggi di corrispondenza e riducibilità psicofisica. È dall'esemplificazione di

185 Kim 1998, p. 116.
186 Cfr. Kim 1993b, pp. 364-65.
187 Dove "localmente" è da riferirsi a "specie-specifiche".

un singolo realizzatore fisico-neurale specie-specifico che, secondo Kim, traiamo per proiezione, l'esemplificazione del corrispondente genere mentale.

In questa visione le proprietà mentali svolgono il ruolo di "intermediari causali" tra le sensazioni e i comportamenti. Tuttavia solo le proprietà fisiche sono in grado di svolgere questo ruolo causale ed è per questo che le proprietà mentali devono avere una realizzazione fisica. Le proprietà mentali sono dunque proprietà del secondo ordine, per avere le quali è necessario che sussistano certe proprietà fisiche, ovvero proprietà del primo ordine:

> *F* is a *second-order property* over a set **B** of base (or first-order) properties iff *F* is the property of having some property *P* in **B** such that *D(P)*, where *D* specifies a condition on members of **B**[188].

Le proprietà dell'insieme **B** sono i realizzatori delle proprietà di secondo ordine; quindi avere uno stato mentale, come per esempio un dolore, può essere descritto sia come avere una certa proprietà mentale, sia come avere la proprietà di stare in un certo stato fisico, ovvero come la proprietà di avere determinate proprietà fisiche che soddisfano certe condizioni.

Sarebbe proprio questa eterogeneità non nomica dei generi mentali a rendere per Kim inattuabile qualsiasi generale "psicologia scientifica". Ciò induce a sostenere una concezione riduzionista in senso locale della psicologia, essendo necessario riconoscere l'esistenza di più teorie psicologiche specie-specifiche a partire dalle singole riduzioni fisiche locali.

2.2.2. La sopravvenienza

Abbiamo già parlato, quando si è vista la critica a Davidson, della visione che Kim ha della sopravvenienza. Questa è per lui un modello descrittivo del rapporto mente-corpo, ma che tuttavia non riesce a spiegare in modo soddisfacente tale rapporto poiché non fornisce alcuna spiegazione effettiva della relazione mente-corpo. Tale modello però, avverte Kim, ha comunque il suo valore, infatti il sostenere una certa versione della sopravvenienza egli lo annovera tra le condizioni minime per considerarsi fisicalisti.

All'idea della sopravvenienza è associata quella di un mondo stratificato in più livelli:

[188] Kim 1998, p. 20.

come livello inferiore viene generalmente assunto quello delle particelle elementari (gli elementi di base della materia), seguono atomi, molecole, cellule, organismi viventi e così via. Dunque, in base a questa visione, entità di un certo livello (escluse ovviamente le più elementari) sono scomponibili, senza residui, in entità del livello immediatamente inferiore. Da ciò ne consegue che alcune proprietà le troviamo solo a certi livelli e sono caratteristiche solamente di quei livelli: «a neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituent neurons, or subassemblies, and human beings have causal powers that none of our individual organs have»[189].

Una proprietà mentale, realizzata da certe proprietà fisiche, occorre necessariamente quando occorrono i suoi realizzatori; dunque una proprietà mentale sopravviene ai suoi realizzatori. Questo tipo di relazione è di tipo nomologico.

Il modello proposto da Kim è quello della causazione sopravveniente, che egli definisce anche "strong supervenience":

> The general model of supervenient causation applied to macro-causal relations is this: macro-event **m** is a cause or effect of event **E** in virtue of the fact that **m** is supervenient on some micro-event, **n**, which is a cause or effect of event **E**. The suggestion then is that we use this model to explain mental causation: a mental event is a cause, or an effect, of another event in virtue of the fact that it is supervenient on some physical event standing in an appropriate causal relation to this event. Thus, mental properties are seen as deriving their causal potential from the physical properties on which they supervene. [...] Supervenient causation requires talk of specific mental properties supervening on specific physical base properties, and this is possible only if there are laws correlating psychological with physical properties[190].

Le proprietà o gli stati mentali sopravvenienti desumono o ereditano l'efficacia causale dalle loro rispettive basi fisiche di sopravvenienza; le mie credenze e volontà possono innescare causalmente una serie di processi o comportamenti a loro volta causalmente efficaci, come l'aumento della temperatura causa il macrofenomeno della bollitura dell'acqua

189 Ivi, p. 85.
190 Kim 1989, pp. 45-46; cfr. anche in Kim 1993b pp. 282-83.

e quindi l'evaporazione.

Dunque per Kim l'esemplificazione di un certo genere mentale in una data occasione, in virtù del suo realizzatore fisico, eredita e conserva i poteri causali del suo realizzatore mediante la relazione di sopravvenienza, la quale diventa compatibile col fisicalismo riduzionista. Le proprietà del secondo ordine non introducono alcuna nuova potenza causale nel mondo, per il semplice fatto che la loro efficacia dipende da quella già presente nel mondo fisico, esibita dalle proprietà del primo ordine e dalle quali la ereditano in base al principio di eredità causale.

M sopravviene a P perché P lo realizza, avere M consiste nell'avere P; avere M è avere P. Sono le leggi che prevalgono nel nostro mondo a determinare che certe proprietà fisiche sono i realizzatori di una data proprietà mentale. Questo tipo di caratterizzazione è per Kim una spiegazione (che era ciò che ci aspettavamo la nozione di sopravvenienza ci fornisse del rapporto mente-corpo), perché la nozione di realizzazione (in accordo con certe leggi vigenti nel mondo) non esprime una mera correlazione, ma riconduce quella correlazione sotto leggi note.

Kim quindi sostiene un'identità tra proprietà; infatti quando si dice, per esempio, che la temperatura di un gas è uguale all'energia cinetica di quel gas, non si intende con ciò affermare che una certa cosa è uguale a un'altra cosa; bensì si intende dire che un suo aspetto o una sua proprietà corrisponde (non sempre ma in determinate circostanze) a un altro aspetto o a un'altra proprietà:

> The items that need to be identified are properties—that is, we would need to identify mental properties with physical properties. If this could be done, that would be an excellent way of vindicating the causal powers of mentality[191].

La strategia riduttiva per identificare una proprietà mentale con i suoi realizzatori fisici è quella della *riduzione funzionale*; ridurre una proprietà o un fenomeno consiste per Kim proprio in questo: ricostruirli funzionalmente in termini di relazioni causali con altre proprietà e fenomeni.

2.2.3. La riduzione funzionale

[191] Kim 1989, p. 45.

A questo punto Kim ci spiega che tipo di riduzione del mentale al fisico egli ha in mente, chiamando questa *riduzione funzionale*, paragonandola a quella tra geni e DNA. La riduzione alla quale si riferisce è diversa dalla riduzione "classica" che si rifà alle leggi-ponte di Nagel. Il modello proposto da Nagel consente di derivare le leggi della teoria che si vuole ridurre dalle leggi della teoria di base attraverso leggi-ponte. Queste sono le stesse leggi che abbiamo visto Davidson negava esserci tra predicati mentali e fisici.

In questo tipo di riduzione Kim identifica un problema importante e fondamentale, cioè il fatto che «by adding the bridge laws to the reductive resources as auxiliary premises, *Nagel reduction extends the reduction base.*[...] In a Nagel reduction the bridge laws connecting the two domains enter the picture as indispensable supplementary assumptions, and this means that the base theory has been expanded both in its ideology (via addition of new predicates/concepts) and its ontology (via addition of new properties»)[192].

Ma, argomenta l'autore, se la riduzione deve essere un processo di spiegazione che produce una spiegazione delle leggi e dei fenomeni che devono essere ridotti sulle basi delle leggi della teoria di base, la riduzione così come intesa da Nagel non produce tale spiegazione. Infatti nelle leggi bicondizionali del tipo M → P, M e P restano proprietà distinte e non vi è quindi nessuna semplificazione, né concettuale, né ontologica. L'unica possibilità consiste per Kim nel passare da M → P a M = P, cioè dalla correlazione all'identificazione delle proprietà mentali con le fisiche. E questo è possibile solo grazie alla riduzione funzionale. Funzionalizzare M significa assumerlo come un designatore non rigido[193], perché M risulta definito nei termini di un'altra proprietà dotata, in maniera contingente, di efficacia causale. Questa efficacia dipende infatti da leggi fisiche valide in questo mondo e la condizione grazie alla quale M viene definita è dunque di tipo nomologico (cioè non necessariamente sussistente in tutti i mondi possibili), quindi M = P risulta contingente.

2.2.4. Tre tappe per l'identificazione delle proprietà mentali

Per ridurre una proprietà, spiega Kim, per esempio essere un gene, bisogna

[192] Kim 2005, pp. 99-100.

[193] Si è parlato dei designatori rigidi quando abbiamo visto la critica di Kripke alla teoria dell'identità di tipo; comunque ricordiamo che questi sono espressioni che identificano l'oggetto a cui si riferiscono, attraverso le sue proprietà essenziali, in tutti i mondi possibili.

innanzitutto funzionalizzare il gene stesso, ovvero bisogna definirlo, o ridefinirlo, in termini dei compiti causali che la proprietà esegue. Dunque potremo definire in un primo momento un gene come un meccanismo che codifica e trasmette informazioni genetiche, successivamente dovremmo trovare i realizzatori della proprietà funzionale che abbiamo descritto (la proprietà di codificare e trasmettere informazioni genetiche) e quindi la proprietà nel dominio di riduzione di base che svolge lo specifico compito causale. Scopriremo così che sono le molecole di DNA i meccanismi che svolgono il compito di codificare e trasmettere le informazioni genetiche. Da ultimo dobbiamo avere una teoria esplicativa che spieghi appunto come i realizzatori della proprietà che è stata ridotta svolgono il compito causale. In questo caso per esempio tale teoria può essere la biologia molecolare. Dunque il modello di riduzione funzionale di Kim consiste nella riduzione di proprietà attraverso la loro identificazione con la proprietà di base, perciò la proprietà funzionalizzata (poniamo M) diventa una proprietà di secondo ordine definita dal suo ruolo causale, ovvero da una specificazione causale (H) che descrive il suo ruolo causale; quindi la proprietà funzionalizzata (M) diventa la proprietà di avere una proprietà con un determinato potenziale causale, e risulta che P è esattamente la proprietà che è adatta alla specificazione causale. Questo tipo di riduzione dunque ci offre una spiegazione riduttiva dei processi che si differenzia in tre fasi:

1. l'interpretazione, o reinterpretazione, delle proprietà che devono essere ridotte in base al ruolo causale che supponiamo esse abbiano;
2. ricerca dei realizzatori fisici delle proprietà funzionali, ovvero i meccanismi o le proprietà che eseguono lo specifico ruolo, o lavoro, causale;
3. identificazione dei ruoli con i realizzatori.

Ciò che è importante per la riduzione funzionale, sottolinea Kim, non è se siamo ancora in grado di identificare i realizzatori, ma che la proprietà possa essere definita in termini del ruolo causale:

> that a property is functionalizable – that is, it can be defined in terms of causal role – is necessary and sufficient for functional reducibility[194].

Facciamo un esempio, prendiamo in considerazione il termine "dolore": innanzitutto dobbiamo isolare il ruolo causale del dolore; poi si devono ricercare i meccanismi che

[194] Ivi, p. 165.

implementano i ruoli precedentemente isolati, e infine si deve identificare la proprietà funzionale isolata con i realizzatori. Quindi secondo Kim per un sistema, poniamo S, avere una proprietà funzionale M significa per questo sistema avere una proprietà di primo ordine P che occupa il ruolo causale definito da M.

Tuttavia Kim sottolinea come la costruzione di proprietà di secondo ordine, una volta date quelle di base, non è un processo che si svolge con un movimento dal basso verso l'alto, dal micro al macro, o viceversa. Proprietà del primo e del secondo ordine sono proprietà dello stesso oggetto, così come la "dormitività" e le proprietà chimiche che la realizzano sono entrambi pertinenti alla stessa pillola che fa dormire. Per esempio, il Valium e il Seconal possiedono entrambi la proprietà funzionale della "dormitività"; questa proprietà di secondo ordine però viene svolta da due diversi realizzatori: il "diazepam" e il "seconbarbital"[195]. Per qualcosa, avere una proprietà funzionale del secondo ordine è, per essa stessa, avere una proprietà del primo ordine che soddisfa certe particolari condizioni, le quali definiscono e specificano quella proprietà funzionale.

A questo punto Kim introduce la differenza, per lui fondamentale, tra ordini e livelli: per *ordine* si intende la gerarchia che esiste tra proprietà funzionali e la loro realizzazione (proprietà di primo ordine e proprietà di secondo ordine); per *livello* invece si intende la gerarchia che si crea nel passaggio dal micro al macro (come per esempio dall'unione di due molecole di idrogeno e una di ossigeno – microlivello - si ottiene l'H_2O – macrolivello).

Le proprietà sopravvenienti e le loro proprietà di base sono istanziate dagli stessi oggetti e quindi sono sullo stesso livello. Proprietà di ordine diverso possono applicarsi a oggetti dello stesso livello. Così, mentre la "dormitività" e le proprietà chimiche che la realizzano si applicano agli stessi oggetti (entrambe si applicano alle pillole, la "dormitività" e le proprietà chimiche sono allo stesso livello, tuttavia sono di ordine diverso, essendo la "dormitività" una proprietà di secondo livello, e le proprietà chimiche proprietà di primo livello), la trasparenza dell'acqua non si applica all'idrogeno e all'ossigeno (perché l'H_2O e le molecole di idrogeno e ossigeno non sono allo stesso livello).

> a second-order property and its realizers are at the same level in the micro-macro hierarchy; they are properties of the very same objects. […] For something to have a second-order property is for it to have one or another of its realizers, that is, a

[195] Kim 1998, pp. 20-21.

first-order property satisfying the specification that defines the second-order property [196].

La sopravvenienza si applica alle relazioni tra ordini che appartengono allo stesso livello, le proprietà di secondo ordine occorrono allo stesso livello di quelle del primo ordine. Questa relazione di sopravvenienza dunque «does not have the effect of emptying macrolevels of causal powers and rendering familiar macro-objects and their properties causally impotent»[197].

La spiegazione riduttiva, piuttosto che la riduzione quindi, può salvaguardare l'autonomia ontologica dei fenomeni non fisici, ovvero quelli di alto livello, ovvero quelli mentali. La spiegazione riduttiva, proprio perché procede per più gradi, secondo Kim «gives us a more realistic picture of reduction in the sciences»[198].

Simone Gozzano ci fa notare che nonostante Kim ritenga che le proprietà mentali siano di secondo ordine, il suo funzionalismo si avvicina a quello che viene denominato "dei realizzatori", che interpreta le proprietà mentali come entità di primo ordine, piuttosto che al funzionalismo dei ruoli, per il quale esse sono proprietà di secondo ordine. Il funzionalismo dei realizzatori, secondo Gozzano, è l'unico che salvaguarda l'efficacia causale del mentale; infatti per il funzionalismo dei ruoli una volta interpretate in termini funzionali le nozioni mentalistiche sono sufficientemente individuate, lasciando aperta la questione ontologica, ovvero di che tipo sono queste entità funzionali. Per Kim invece «sui risultati delle tecniche di individuazione del funzionalismo, vanno superimposte delle conseguenze ontologiche, le quali sanciscono che tali risultati sono stati e proprietà fisiche, e non proprietà astratte da realizzare in questo o quel modo»[199], avvicinando così la sua posizione a quella del funzionalismo dei realizzatori.

Kim quindi afferma che stati quali credenze e desideri sono definibili in termini funzionali, e per spiegare meglio le sue motivazioni fa un esempio: supponiamo che esista una popolazione che sia nelle funzioni e nel comportamento indistinguibile dalla nostra, interagiscono con l'ambiente fisico e tra di loro proprio come facciamo noi; questa popolazione mostra anche dei comportamenti linguistici analoghi ai nostri, e quindi sono

[196] Ivi, p. 82.
[197] Ivi, p. 86.
[198] Kim 2005, p. 102.
[199] Gozzano 2007, pp. 168-69.

creature che usano un linguaggio. Se così stanno effettivamente le cose, prosegue Kim, è coerente attribuire a tale popolazione stati quali credenze, desideri, intenzioni, poiché esprimere un enunciato corrisponde a esprimere una credenza e quindi, conclude Kim, un essere dotato di linguaggio è un essere cognitivo e intenzionale. A questo punto è chiaro per l'autore come gli stati intenzionali (come appunto un pensiero, un desiderio, ecc.) sopravvengano sui fatti fisici. Il fatto che queste creature mostrino un comportamento e un funzionamento simile al nostro ci porta a pensare a loro come creature dotate di una mentalità fortemente analoga alla nostra. E questa è una ragione molto convincente per pensare che «such mental properties are definable and interpretable in terms of their roles in behavior causation. […] That is because these states and processes, like perception, belief, memory, and using information in guide action, are specifiable in terms of their causal roles, or "job descriptions". A creature, or system, that has the capacity to do certain things in certain ways under certain conditions is ipso facto something that perceives, remembers, and appropriately behaves»[200].

Kim ritiene dunque che il comportamento sia strettamente connesso con le proprietà mentali intenzionali e che un'analisi, anche parziale, di queste proprietà in termini del ruolo causale che svolgono, ci aiuta a capire quali sono i meccanismi fisici sottostanti. Finché gli stati intenzionali sono all'interno del dominio del comportamento e includono meccanismi fisici nella loro realizzazione essi non ci portano fuori dal dominio fisico.

Tuttavia, come anticipato prima, lo stesso discorso non vale per gli stati *qualitativi*, infatti «qualia are not functional properties»[201]; non possiamo distinguere, sostiene Kim, il dolore dal prurito, e altre sensazioni, sulla base del loro ruolo causale.

Questo porta Kim a concludere che un fisicalismo globale non è possibile poiché i *qualia* sono i "residui mentali" che non possono rientrare nel dominio fisico:

> qualia, therefore, are the "mental residue" that cannot be accommodated within the physical domain. This means that global physicalism is untenable[202].

Questo "residuo mentale" comprende solo gli stati qualitativi e non tocca gli stati intenzionali. I qualia stanno fuori dal dominio fisico e per questo non hanno nessun ruolo

200 Kim 2005, p. 166.
201 Ivi, p. 168.
202 Ivi, p. 170.

causale, ma questo non fa differenza nell'individuazione delle funzioni degli stati intenzionali. Le similitudini e differenze tra i qualia possono essere salvate, ma ciò che non possiamo salvare, afferma l'autore, sono le qualità intrinseche, per esempio il fatto che il giallo sembri tale, che una rosa profumi in un determinato modo. Mentre i qualia in sé non possono essere caratterizzati in termini funzionali, le differenze tra i qualia possono avere questo tipo di caratterizzazione. Non fa differenza in termini pratici se, per esempio, l'esperienza del rosso o del verde è invertita: noi dovremmo ancora essere capaci di interpretare in modo corretto, per esempio, un semaforo.

Kim ritiene così di aver risolto, anche se non completamente, il problema della causazione mentale. Per tutte queste motivazioni Kim ritiene che «physicalism is not the whole truth, but it is the truth near enough, and near enough should be good enough»[203].

Tuttavia, come ci fa notare Gozzano, pare difficile che una proprietà mentale sia sempre ed esclusivamente o intenzionale o fenomenica: pensiamo a un'emozione dotata di contenuto intenzionale, dove dovremmo inquadrarla nel modello proposto da Kim?[204] Come possiamo, tra l'altro, avere più di una sostanza fondamentale senza averne due? Vediamo alcune critiche che sono state mosse nei confronti della posizione di Kim.

2.3. Alcune critiche

Come abbiamo visto Kim, contro la teoria della realizzabilità multipla, propone un modello di riduzione in base alla specie, dove uno stato mentale M ha diversi realizzatori in differenti specie o strutture, quindi ciò che abbiamo è un'identificazione ridotta e ristretta nella quale una proprietà mentale o di secondo ordine M per un sistema S preso in considerazione è uguale a P, ovvero a una proprietà fisica o di primo ordine, per ogni genere di S nel quale M può essere realizzato.

Secondo Ausonio Marras[205] questo tipo di riduzione non si differenzia molto dal modello proposto da Nagel, infatti entrambi porterebbero all'isomorfismo tra i due livelli di leggi coinvolte nella riduzione e la conseguente postulazione di leggi bicondizionali che legano i due livelli. Il modello di Kim, argomenta Marras, insiste sulla riduzione ristretta ai generi, la quale comprende l'identificazione teoretica delle proprietà funzionalizzate con i loro

203 Ivi, p. 174.
204 Per un'accurata disamina di questo punto cfr. Gozzano 2007, pp. 202-203.
205 Cfr. Marrar 2002.

realizzatori specifici; questo modo sarebbe equivalente alle leggi ponte bicondizionali «mediating the derivation of the higher-level laws that articulate the "role" of the functionalized properties in the target theory, from the lower-level laws conformity to which enables the role-filler properties in the base to fill their respective roles»[206].

Tra l'altro, contrariamente a quanto sostenuto da Kim, per Marras non c'è ampliamento della teoria di base, né cambiamento di questa, l'uso di leggi ponte non implica l'espansione dell'ideologia e dell'ontologia della teoria di base. Una riduzione di successo garantisce una reinterpretazione delle leggi ponte come esprimenti "identità teoretiche", portando a una semplificazione ontologica piuttosto che a un'espansione ontologica, come sostiene Kim. Dovremmo dire altrimenti che la teoria da ridurre è stata ridotta a una teoria ibrida che contiene sia la teoria di base che le leggi ponte, ma questo è scorretto[207].

Per Terence Horgan[208] invece la posizione di Kim è un materialismo dualistico: ripudiando la teoria dell'identità di tipo egli adotta un'ontologia che vede le proprietà come multistrati. Kim ritiene che le proprietà mentali sono determinate da, o sopravvenienti a, le sottostanti proprietà fisiche. Se a questo uniamo il principio di chiusura causale adottato da Kim, del quale abbiamo parlato e che, lo ricordiamo, afferma che ogni stato o evento fisico, quando descritto in termini fisici, è interamente spiegabile in base a leggi fisiche, per Horgan arriviamo a un tipo di dualismo delle proprietà, e non solo: «it is a dualistic materialism—a materialism with a stratified, multi-level, property-ontology»[209].

Più specificatamente un materialismo riduttivo dualista, poiché oltre a negare che le proprietà mentali sono uguali alle proprietà fisiche (e quindi abbracciando una posizione dualista), sostiene anche una posizione materialista che richiede che la psicologia sia riducibile alle scienze naturali. La posizione di Kim abbraccia la chiusura causale del mondo fisico, il dualismo delle proprietà e il fatto che le proprietà mentali siano proprietà causali, e per questo Horgan ritiene che sia implicato nel compatibilismo causale, ovvero nel fatto che la causazione mentale attraverso le proprietà non fisiche può coesistere con la causazione fisica anche se il dominio fisico è causalmente chiuso, e quindi nella causazione verso il basso, cioè quella che va dal mentale al fisico. Tutte queste premesse, dice Horgan, sono presupposte anche dalle posizioni non riduzioniste che Kim pretende negare.

206 Ivi, p. 251.
207 Cfr. ivi, p. 235.
208 Cfr. Horgan 1996.
209 Ivi, p. 581.

Un altro problema nella posizione di Kim, Horgan lo identifica nelle proprietà disgiuntive, le quali non sono proprietà causali, dice Horgan; infatti ogni volta che una tale proprietà viene istanziata, il "lavoro causale" sarà fatto non dall'intera proprietà, bensì solo da una qualunque specifica disgiunzione che è istanziata nella data occasione[210].

Per Horgan il paragone tra la giada e i termini mentali non regge perché i generi mentali non sembrano essere implicati nell'unicità della realizzabilità fisica di generi mentali negli esseri umani. I concetti mentali sono diversi dal concetto di giada. Questo, essendo un concetto di genere fisico, comporta il fatto che la giada è un genere fisico singolo con un unico tipo di composizione fisico-chimica. Tale argomento dunque non impensierisce i sostenitori dell'antiriduzionismo poiché i concetti o generi di alto livello non sono vincolati dal fatto che i generi che loro specificano sono localmente riducibili, mentre la giada pretende di individuare (o cogliere) un solo tipo di causalità fisica.

Paternoster[211] inoltre ci fa notare come la relativizzazione dell'identità alle diverse specie permette comunque l'obiezione che, in una data specie, occorrenze diverse di uno stesso tipo mentale M possono essere realizzate in modo fisicamente diverso. Il modello proposto da Kim porrebbe un vincolo molto stretto, ovvero quello di dovermi trovare nello stesso stato fisico ogni volta che credo che P.

Contro la posizione di Kim è stata avanzata anche un'altra critica che va sotto il nome di "argomento generalizzato"[212], la quale pretende mostrare che, se la visione di Kim funziona, questa condanna all'impotenza causale tutte le proprietà tranne, forse, quelle fisiche. Se, in generale, le proprietà delle scienze speciali sopravvengono sulle proprietà fisiche di base, l'argomento della sopravvenienza si deve applicare alle proprietà di tutte le scienze con l'esclusione della fisica di base. Se non c'è problema ad ammettere la causazione chimica o biologica perché, si chiedono i critici di Kim, tale problema dovrebbe sorgere per le proprietà mentali?

Per difendersi da questa accusa Kim ha proposto la distinzione, della quale abbiamo parlato in precedenza, tra livelli e ordini, spiegando che l'argomento della sopravvenienza si applica solo alle proprietà di secondo ordine, ovvero alle proprietà mentali, e non alle proprietà delle scienze speciali in generale, che sono dislocate a diversi livelli. L'argomento generalizzato assume che una relazione di sopravvenienza che si applica alle proprietà di

210 Ivi, p. 595.
211 Cfr. Paternoster 2002, p. 67.
212 Cfr. Gozzano 2008.

secondo ordine e ai rispettivi realizzatori di primo ordine, valga per proprietà ordinate dalla relazione micro-macro, che è una relazione tra livelli. L'illusione che la sopravvenienza si applichi alle gerarchie macro-micro è generata dal fatto che in molti casi i realizzatori di primo ordine delle macroproprietà di secondo ordine sono proprietà microstrutturali, ovvero proprietà che i macrooggetti hanno in virtù della loro composizione mereologica. Ma queste proprietà microstrutturali non sono microproprietà, piuttosto sono loro stesse macroproprietà che occorrono allo stesso livello delle macroproprietà che essi realizzano. Dunque la relazione di sopravvenienza è una relazione tra livelli (*intralevel*): proprietà che occorrono a differenti livelli lungo la gerarchia micro-macro non sono collegate dalla sopravvenienza, e quindi l'argomento della sopravvenienza non può essere generalizzato a loro.

Gozzano a questo punto si chiede se Kim riconosce due classi di proprietà, «quelle per le quali le identità valgono nomologicamente *e* metafisicamente e quelle per le quali valgono solo nomologicamente»[213]. Tale caratterizzazione porterebbe per Gozzano o al fallimento della sopravvenienza forte, conseguenza che abbiamo visto essere per Kim inaccettabile, o al dualismo delle proprietà, scenario altrettanto inaccettabile per Kim.

CAPITOLO 3

[213] Gozzano 2007, p. 168.

Hilary Putnam: la sopravvenienza globale

3.1. La posizione di Putnam

Vediamo ora come Putnam si inserisce nella disputa Kim-Davidson e quali sono le sue riflessioni riguardo la questione delle proprietà mentali e della sopravvenienza.

L'idea di Putnam si discosta dalle due precedenti prese in esame e questo perché, influenzato da pensatori quali William James, John Dewey, John Austin e Wittgenstein, l'autore difende una posizione secondo la quale «the mind is not a thing; talk of our minds is talk of world-involving capabilities that we have and activities that we engage in»[214]. E ancora

> mind talk is not talk about an immaterial part of us but rather a way of describing the exercise of certain abilities we possess, abilities that supervene upon the activities of our brains and upon all our various transactions with the environment but that do not have to be reductively explained using the vocabulary of physics and biology, or even the vocabulary of computer science[215].

Notiamo da subito quindi come Putnam critica l'idea secondo la quale la mente è qualcosa che sta nelle nostre teste, o più precisamente dentro le nostre teste, e che per questo ci sia bisogno di un modo per capire come identificarla e come questa si relazioni con il corpo e il mondo.

Secondo l'autore quest'immagine della mente collocata all'interno della testa degli

[214] Putnam 1999, pp. 169-70.
[215] Ivi, pp. 37-8; cfr. anche p. 44.

esseri umani è causata da un'idea di percezione, affermatasi con Cartesio e poi ereditata dal pensiero moderno, in base alla quale ci sarebbe un'interfaccia cognitiva, un'entità intermedia, costituita dai dati di senso, tra le menti e il mondo, tra il soggetto e l'oggetto.

La fonte dei nostri dilemmi sulla natura della mente sarebbe perciò causata dall'emergere della concezione della mente secondo la quale la categoria dei fenomeni mentali che riguardano il mondo sono resi possibili da dati sensoriali (o sensazioni)[216], e questi sono concepiti come immateriali:

> the mind/body problem has become (among English-speaking philosophers) the problem of the relation of these apparently immaterial sensations (now thought of as paradigm of the "mental") to the physical world[217].

Un chiaro esempio, secondo Putnam, delle conseguenze di tale concezione della mente è dato proprio dall'argomento di Kim, argomento che presuppone che le proprietà mentali sono proprietà "interne" al nostro corpo e che per questo motivo ci sia bisogno di trovare un modo per spiegare come proprietà mentali e proprietà fisiche siano correlate l'una con l'altra. Ma procediamo con ordine e vediamo i punti salienti dell'argomentazione di Putnam.

3.2. La critica a Kim

Come detto nei capitoli precedenti, Kim ha messo in discussione il fatto che nel monismo anomalo le proprietà mentali siano causalmente efficaci, concludendo che sembrano essere epifenomeni. Lo ricordiamo, per Kim il monismo anomalo implica che, all'interno di un mondo, si darebbe la stessa rete di relazioni causali se si ridistribuissero a piacere le proprietà mentali sui suoi eventi; non verrebbe disturbata alcuna relazione causale particolare se si riassegnassero le proprietà mentali agli eventi a caso e in modo arbitrario, o addirittura se si rimuovesse del tutto la mente dal mondo. Secondo il monismo anomalo, conclude Kim, gli eventi sono cause o effetti solo nella misura in cui esemplificano leggi fisiche, e ciò significa che le proprietà mentali di un evento non fanno alcuna differenza causale. In breve Kim, seguendo questa linea di ragionamento, arriva a proporre una disgiunzione tra due ipotesi: o

[216] Cfr. Putnam 1994, p. xvii.
[217] Ivi, p. xix.

accettiamo l'esistenza di leggi ponte tra mentale e fisico (salvaguardando così l'efficacia causale del mentale), o le proprietà mentali sono epifenomeni.

Putnam parte analizzando l'ultimo corno della disgiunzione, ovvero la possibilità che le proprietà mentali possano essere epifenomeni. Egli sostiene molto esplicitamente che l'epifenomenismo è «C-R-A-Z-Y»[218]; anche Kim non accetta come plausibile l'epifenomenismo, ma per ragioni diverse rispetto a quelle di Putnam.

Le nostre decisioni, argomenta Putnam, possono essere cause di eventi fisici, e sostenere che le decisioni sono totalmente inefficaci significa per l'autore «to reject ways of talking that are essential to our whole conception of ourselves as beings in a world»[219]. Riportiamo l'esempio fornito dallo stesso Putnam: supponiamo che mi voglia fare un bagno e che quindi stia facendo scorrere l'acqua per riempire la vasca. Se qualcuno mi chiedesse "Perché sta scorrendo l'acqua nella vasca?" la mia risposta potrebbe essere "Sto facendo scorrere l'acqua nella vasca per farmi un bagno", ma potrei anche benissimo rispondere "Ho deciso di fare un bagno", e nessuno dubiterebbe che non ho spiegato perché l'acqua sta scorrendo nella vasca. Quindi se il "perché" della risposta "Ho deciso di fare un bagno" è corretto, ovvero risponde in maniera corretta alla domanda "Perché sta scorrendo l'acqua nella vasca?", allora in un certo senso di "causa", argomenta Putnam, la mia decisione di farmi il bagno è la causa del mio aprire il rubinetto in una determinata posizione e dunque dello scorrere dell'acqua nella vasca. Supporre che questo "perché" non sia corretto, ovvero affermare che non è stato a causa della mia decisione di farmi il bagno che l'acqua a quella determinata temperatura sta scorrendo nella vasca, e che quindi la mia decisione è stata totalmente inefficace significa per Putnam, come detto prima, rifiutare modi di parlare che sono essenziali alla concezione che abbiamo di noi stessi come esseri in un mondo.

Putnam dunque sostiene che si possa parlare di eventi fisici come quello appena descritto ("l'acqua che scorre") come aventi sia una spiegazione fisica (es. l'acqua scorre perché ho aperto il rubinetto) sia una spiegazione mentale (es. l'acqua scorre perché ho deciso di farmi il bagno). O, volendola mettere in altri termini, potremmo chiedere "Qual è la causa dello scorrere dell'acqua?" e una risposta del tipo "L'acqua scorre a causa del mio desiderio di farmi il bagno" ha per Putnam lo stesso valore esplicativo di una risposta del tipo "L'acqua scorre a causa del fatto che ho aperto il rubinetto".

[218] Putnam 1999, p. 78.
[219] Ivi, p. 77.

Secondo questa visione, quindi, ci sono diverse risposte alla domanda "Perché sta scorrendo l'acqua?", incluse risposte che si riferiscono alle decisioni, ai pensieri e ad altri fenomeni mentali. Questi ultimi per Putnam non possono essere ridotti alla fisica del cervello per ragioni esterniste, cioè per la dipendenza dei nostri pensieri, desideri, ecc., da ciò che sta fuori dal nostro organismo:

> Part of the notion of a decision is that decisions can be *acted* upon, they can be *carried out*, and my decision to run that water, plus the fact that I carried it out, really did *bring it about* that water is running now[220].

Una spiegazione delle cause degli eventi fisici a livello di stati mentali quali credenze e desideri, è dunque per Putnam assolutamente valida. A sua volta, la nozione di causa è per l'autore fortemente legata al soggetto, ai suoi interessi, oltre che al concetto di spiegazione. Vediamo di cosa si tratta.

3.2.1. La causazione mentale

Come si evince dall'esempio precedente, la verità di un giudizio nella forma "X causa Y" dipende, per Putnam, dal contesto e dagli interessi delle persone che esprimono il giudizio[221], quindi, per esempio, da cosa il parlante vuole sapere in un contesto particolare; spiegazioni differenti generalizzano differenti classi di casi.

La nozione di causazione è interdipendente rispetto alla nozione di spiegazione: «the notions of explanation and causation presuppose one another at every point; neither has "priority" in the sense of being reducible to the other»[222]; e ancora «causation and explanation are interdependent notions»[223]. La causazione, argomenta Putnam, esiste solo in quello che è il nostro mondo della vita, il mondo in cui vivono gli esseri umani, ciò che la rende fenomenologicamente "reale" è la possibilità di controllare le asserzioni causali, di vedere come nuovi dati e nuove teorie si applicano a esse, per questo afferma che: «the ordinary

220 Ivi, p. 146.
221 Cfr. Putnam 1992, p. 64.
222 Putnam 1999, p. 137.
223 Ivi, p. 145.

notion of causality is a *cognitive* notion»[224]. Il mondo in cui viviamo (il mondo descritto dal linguaggio ordinario) è pieno di cause ed effetti, ma non impone un'unica totalità delle possibili descrizioni una volta per tutte. E questo perché, nella visione dell'autore, la realtà non è una singola supercosa, ma è data dai modi con i quali «we endlessly renegotiate – and are *forced* to renegotiate – our notion of reality as our language and our life develop»[225].

La descrizione della realtà dunque non è mai mera copia della stessa, anzi ampliamo costantemente i modi in cui il linguaggio risponde alla realtà e parimenti ampliamo i nostri concetti di causazione. Citando John Aldane, Putnam si dice d'accordo con l'affermazione che ci sono «as many kind of cause as there are senses of 'because'», ma precisa, non nel senso che i diversi sensi di "perché" sono significati diversi della parola. Quando parliamo di diversi "sensi" di una parola o di una frase, argomenta Putnam, ciò che facciamo è indicare i diversi usi dell'espressione in questione, usi che sono proiezioni di un singolo concetto. La frase di Haldane, prosegue Putnam, è da intendersi come il fatto che capiamo i vari usi di "causa" padroneggiando le varie pratiche esplicative.

Ritornando all'esempio precedente dell'acqua che scorre nella vasca, non c'è nessuna contraddizione nel descrivere questo evento come "aver girato il rubinetto in una certa posizione" e "aver deciso di fare il bagno", entrambe sono cause dello stesso evento e spiegazioni dello stesso evento. Non si deve credere, argomenta Putnam, nell'esistenza di un'unica visione causale che sia migliore di tutte le altre; ciò che abbiamo sono invece visioni migliori o peggiori, ed è proprio quest'ultima possibilità a conferire un contenuto alla nozione di oggettività[226]. Seguendo Wittgenstein, Putnam quindi sostiene che non esiste un unico modo per descrivere la realtà e che, parimenti, non esiste un unico linguaggio vero e significante per esprimere i fatti del mondo.

Per questo egli sostiene che le affermazioni che riguardano credenze, desideri, ecc., ovvero quelle affermazioni che generalmente vengono considerate "inferiori" dai materialisti rispetto a quelle della fisica, hanno invece un valore di autenticità: riprendendo le parole di James Conant, Putnam ritiene che le affermazioni sul mentale sono governate da norme di verità e validità come qualsiasi altra affermazione.

Dunque riassumendo, per Putnam le proprietà mentali non possono essere epifenomeni in quanto le caratteristiche rilevanti nella causazione riguardano il *nostro* modo

[224] Putnam 1992, p. 78.
[225] Putnam 1999, p. 9.
[226] Cfr. Putnam 1987, p. 77.

di spiegare perché l'evento X causa l'evento Y, e la mera relazione causale non può fare il lavoro di una relazione cognitiva. Quindi "causa" (nel senso ordinario del termine) può essere parafrasato da una locuzione che coinvolge "spiegare": per esempio, possiamo dire che la presenza nell'ambiente di certi corpi con composizione chimica H_2O "causano" il nostro uso della parola "acqua", ma possiamo anche dire che la presenza di H_2O nel nostro ambiente, la nostra dipendenza da H_2O per la nostra sopravvivenza, ecc., sono parte della spiegazione del nostro avere una parola che usiamo come usiamo la parola "acqua". Ma questa nozione di "causa" non è certo definibile fisicamente. Che cosa è (o non è) una "causa" o una "spiegazione" dipende dalle facoltà conoscitive del soggetto e dalla ragione che lo ha spinto a porre una determinata domanda, a chiedere il perché di qualcosa, il suo interesse relativamente all'evento in questione. E poiché ragioni e interessi mutano col tempo e con le occasioni, «the form of all knowledge claims and the ways in which they are responsible to reality are [not] fixed once and for all in advance».[227] Tra l'altro, sottolinea Putnam, l'interesse è un attributo del pensiero e del ragionamento, non della natura[228].

Questa visione implica che il principio della chiusura causale del mondo fisico, visto come principio ontologico, è falso per Putnam, contrariamente a quanto abbiamo visto sostenere da Kim. Questo è un punto di distacco rilevante tra i due autori, soprattutto se consideriamo che parte della critica di Kim nei confronti di Davidson si basa proprio sulla tesi che il monismo anomalo violerebbe tale principio. Kim, come abbiamo illustrato nel capitolo precedente, è legato a una nozione di causa e di fisica differente rispetto a quella di Putnam.

Quanto detto finora infatti rispecchia anche l'idea che Putnam ha della fisica stessa, cioè come qualcosa di non fisso, di non già dato in maniera definitiva: la scienza stessa, e non solo il linguaggio ordinario, è profondamente pluralistica nella propria ontologia[229]. Kim invece, argomenta Putnam, come altri materialisti, crede nella falsa idea che solo la scienza ci dice come sono le cose veramente. Ma Putnam è di tutt'altro avviso: egli ritiene che non esiste la possibilità di un fondamento ultimo della conoscenza scientifica, così come di ogni altro tipo di conoscenza[230]. Il ché non significa che i filosofi non si devono curare delle scoperte e dei progressi scientifici riguardo percezione e cognizione, anzi «progress with regard to them

227 Putnam 1999, p. 20.
228 Cfr. Putnam 1983, pp. 213-14.
229 Cfr. Putnam 1990, tr. it., p. 227.
230 Cfr. Puntam 1987, p. 79.

requires constant attention to those findings»[231]; tuttavia, quando in discussione ci sono gli esseri umani e le loro capacità, gli studiosi ai quali dovremmo riferirci per capire come caratterizzare tali capacità, conclude Putnam, sono i filosofi del linguaggio, non i fisici.

3.2.2. L'ipotesi degli automi: una condizione intellegibile?

Dopodiché Putnam passa ad esaminare l'altra parte della disgiunzione proposta da Kim, ovvero il fatto che se vogliamo conservare l'efficacia causale del mentale, senza appellarci al dualismo, dobbiamo "arrenderci" al riduzionismo.

Kim sostiene che, nel caso la posizione di Davidson sia corretta, vale la seguente tesi: "(AUTOMI): *Anche nel caso in cui certe persone non avessero proprietà mentali, fintanto che tutte le loro proprietà fisiche e i loro ambienti fisici rimanessero identici, accadrebbero gli stessi eventi fisici*".

Il problema è che, secondo Putnam, questa tesi si basa su di un antecedente la cui intelligibilità è opinabile. Tale antecedente è l'ipotesi da lui chiamata "(SENZ'ANIMA): *Certe persone non hanno proprietà mentali, ma tutte le loro proprietà fisiche sono le stesse che avrebbero nel caso in cui avessero proprietà mentali e i loro ambienti fisici sono gli stessi*".

Qua Putnam fondamentalmente discute la plausibilità dell'esistenza degli *zombie*[232], egli ritiene che questa sia una condizione completamente inintelligibile perché siamo totalmente incapaci di dire come potrebbe risultare possibile che l'ipotesi (SENZ'ANIMA) sia ammissibile. Non si riesce a comprendere come qualcosa possa avere la stessa struttura fisica di un essere umano e non avere coscienza: «to think that zombie are a possibility, I would have to assume that my entire view of the world is wrong in a way I am utterly unable to make cohere with anything I believe»[233]. Vediamo perché Putnam arriva a sostenere questo.

3.2.3. La relazione tra proprietà mentali e proprietà fisiche: indipendenza e identità

[231] Putnam 2012, cap. 3 *Corresponding with Reality*, p. 90.

[232] Abbiamo parlato dell'esistenza degli zombie nel primo capitolo, in riferimento a Chalmers, il quale ha elaborato l'argomento della possibilità logica degli zombie, in evidente contrasto con quanto sostenuto qui da Putnam.

[233] Putnam 2012, cap. 28 *Wittgenstein: a Reappraisal*, p. 487.

Supporre che l'ipotesi (SENZ'ANIMA) sia intellegibile, argomenta Putnam, implica che le proprietà mentali possano essere indipendenti da quelle fisiche: solo così può essere logicamente possibile che le proprietà fisiche possano essere presenti senza le proprietà mentali. Ma questa supposizione ha senso soltanto se ha anche senso supporre che le proprietà mentali sono riducibili alle proprietà fisiche, nel senso di essere nomologicamente coestensive alle proprietà fisiche. Ma la riducibilità psicofisica e l'identità sono loro stesse nozioni inintelligibili: «*the notion of identity has not been given any sense in this context*»[234].

Identificare le proprietà mentali con quelle fisiche, o quelle funzionali, dovrebbe servire a permetterci di spiegare fenomeni (in questo caso quelli mentali) che altrimenti non saremmo in grado di spiegare. Esattamente come la temperatura si è empiricamente rivelata essere identica a l'energia cinetica molecolare media, nello stesso modo si vorrebbero ridurre e identificare le proprietà mentali con quelle fisiche.

Il problema centrale è però che la nozione di stato mentale non è mai stata ben definita: non possediamo modi né per individuare gli stati mentali, né una serie di leggi psicofisiche (non ridotte) che spieghi cos'è uno stato mentale. Ma l'idea di una identificazione teoretica presuppone che i concetti da ridurre (ovvero quelli di stati mentali) siano già di fatto sotto qualche genere di controllo scientifico. Introdurre un insieme di concetti che al momento non figura in nessuna legge e poi iniziare a parlare di ricerca per l'identificazione teoretica di questi stati mentali con gli stati funzionali del cervello è impegnarsi in una identificazione teoretica fantascientifica[235]: «No serious scientifically discussable issue of identity has really been raised».[236]

Infatti, argomenta Putnam, in questo caso non possiamo usare il classico modello di riduzione della termodinamica alla meccanica statistica. Qui sia la teoria ridotta sia quella riducente hanno un corpus di leggi ben definito. Lo stesso discorso non regge per la psicologia: una proprietà mentale non può essere ridotta a una proprietà funzionale poiché la nozione di proprietà funzionale dipende dal formalismo in cui è scritto il programma, ma «no one has the slightest idea what a formalism in which one can write programs that reduce folk

234 Putnam 1999, p. 85. Il rifiuto da parte di Putnam di qualsiasi forma di riduzione dovrebbe ormai essere chiaro, anche alla luce di quanto è stato detto nei capitoli precedenti riguardo alla sua critica nei confronti della teoria dell'identità dei tipi e alle motivazioni che lo hanno spinto ad allontanarsi dal funzionalismo. Tuttavia, dato che qui si sta prendendo in considerazione la critica a Kim, e dato che quest'ultimo propone la *riduzione funzionale* quale metodo per identificare le proprietà mentali, ci pare opportuno ribadire l'idea di Putnam al riguardo.

235 Cfr. Putnam 1997a, p. 37.

236 Putnam 1999, p. 86.

psychology might look like»[237]. Le proprietà funzionali sono definite usando la nozione di causa ed effetto, ma la causazione, come detto in precedenza, non è una relazione unicamente fisica bensì, ribadisce Putnam, una nozione relativa all'interesse[238]. I termini riguardanti gli stati mentali non appartengono alla fisica, non c'è nessuna definizione indipendente di stato mentale che possa essere il soggetto di una teorizzazione psicologica, «no actual psychological theory has ever pretended to provide a set of laws that distinguish, say, the state of being jealous of Desdemona's fancied regard for Cassio from every other actual or possible propositional attitude»[239]. L'idea che ci sia un problema scientifico sulla natura della mente presuppone l'immagine, sbagliata, della mente e dei pensieri come oggetti.

Putnam dunque ritiene che non ci possa essere uno stato mentale identificabile con un determinato stato fisico o un determinato stato funzionale, proprio perché lo stato mentale non è qualcosa di oggettivo che sta fisso lì, presumibilmente nel cervello, e bisogna solo capire cos'è e come caratterizzarlo. Per questo parlare di identità tra stati mentali e stati fisici o stati funzionali per Putnam non ha senso.

3.3. L'esternismo semantico

L'apporto dell'ambiente esterno quando parliamo di stati mentali e di linguaggio è fondamentale per Putnam. Questa visione non è nuova del pensiero putnamiano, infatti comparve già verso la prima metà degli anni '70 quando Putnam formulò il famoso esperimento della Terra Gemella[240]. Egli elaborò tale esperimento nel contesto dell'esposizione della sua concezione causale del significato dei termini di genere naturale e di grandezza fisica, ma è significativo anche dell'attuale pensiero dell'autore riguardo al problema mente-corpo e al discorso sulla sopravvenienza.

L'ipotesi fantascientifica di Putnam ci invita a immaginare che esista un pianeta – la Terra Gemella – completamente uguale alla Terra tranne che per un elemento: l'acqua. Questa infatti nella Terra Gemella, pur avendo lo stesso aspetto esteriore e pur venendo utilizzata nello stesso modo in cui si usa sulla Terra, avrebbe una composizione chimica diversa, indicata da Putnam con la formula XYZ.

---

237 Ivi, pp. 85-6.
238 Cfr. Putnam 1983, pp. 224-25.
239 Putnam 1997a, p 35.
240 Cfr. Putnam 1975, cap. 12 *The Meaning of 'Meaning'*, pp. 223-27.

Se degli astronauti della Terra giungessero sulla Terra Gemella, inizialmente penserebbero che il termine "acqua" abbia lo stesso significato in entrambe le terre, e quindi sia il terrestre che il gemello di Terra Gemella si troverebbero nello stesso stato psicologico rispetto al termine "acqua". Ma dopo una serie di analisi chimiche si renderebbero conto che ciò che in apparenza ha tutte le caratteristiche dell'acqua non è in realtà H_2O, bensì XYZ. Dunque si scopre che sulla Terra Gemella il significato di "acqua" non è H_2O, bensì XYZ, e questo vuol dire che il termine ha estensioni[241] diverse a seconda della Terra considerata[242].

Ciò che Putnam sottolinea è che, finché i due abitanti non scoprono la diversa composizione chimica della parola "acqua", essi condividono lo stesso stato psicologico, nonostante l'estensione del termine "acqua" sia diversa. Dunque "acqua" aveva un riferimento diverso anche prima di tale scoperta, poiché le sostanze erano diverse, ma le rappresentazioni mentali, gli stati psicologici, non erano differenti. Putnam conclude quindi che l'estensione di un termine non è funzione del solo stato psicologico del parlante, ma è parzialmente fissata dall'ambiente stesso: questo è ciò che Putnam definisce il contributo dell'ambiente. Il significato dunque non è qualcosa che appartiene al singolo parlante, bensì all'intera comunità dei parlanti per il semplice fatto di vivere in un determinato contesto: «l'ambiente stesso gioca un ruolo nel determinare a che cosa si riferiscono le parole di un parlante o di una comunità»[243].

La differente estensione dei due termini in questione non dipende dai concetti che si hanno; quindi, conclude Putnam, «meanings just ain't in the *head!*»[244].

L'esempio della Terra Gemella mostra come il significato si correla all'oggetto, in questo caso l'acqua, in virtù di un'operazione sociale:

> What gives one's words the particular meaning they have is not just the state of one's brain, but the relations one has to both one's non-human environment and to other speakers[245].

Lo stesso discorso vale per credenze e desideri e qualsiasi altro stato mentale. Il contenuto di una credenza, avverte Putnam, non è solamente mentale in quanto è possibile, a

241 Ovvero l'insieme delle cose delle quali il termine è vero.
242 H_2O sulla Terra e XYZ sulla Terra Gemella.
243 Putnam 1988, tr. it., p. 60.
244 Putnam 1975, p. 227.
245 Putnam 1997a, p. 196.

parità di stati cerebrali, intrattenere credenze diverse, aventi cioè diverso contenuto, in contesti ambientali differenti[246]. Tra l'altro egli sostiene che non tutti gli esseri umani che hanno lo stesso pensiero (in diverse culture, o con differenti basi di conoscenza e diverse risorse concettuali) hanno in comune una caratteristica fisica e computazionale che possa essere "identificata con" quel pensiero. Il pensiero infatti non è riducibile al livello fisico o funzionale.

Putnam dunque sostiene una posizione secondo la quale la connessione tra il mondo e i nostri pensieri è considerata necessaria perché i fattori esterni al soggetto contribuiscono in modo essenziale alla determinazione dei contenuti mentali. Come per i nomi di genere naturale, Putnam sostiene che dall'esame dei termini di sensazioni non si può concludere che i loro riferimenti siano stati cerebrali: infatti nessuna teoria importante sulla natura della mente può essere confermata o respinta mediante un esame del significato dei termini mentali.

Queste riflessioni sul significato sono importanti poiché a tutt'oggi Putnam crede che «different philosophical pictures of how language functions and what meanings are (or, better, what knowledge of meanings consist in) will affect our attitudes toward just about every philosophical debate»[247]; difficile trovare un'affermazione più chiara sul ruolo fondamentale giocato dal linguaggio e dalla riflessione su di esso nel rendere possibile le nostre credenze su una vasta gamma di questioni.

3.4. Ulteriori riflessioni sull'ipotesi (SENZ'ANIMA)

Ma torniamo alla critica nei confronti di Kim e vediamo come le riflessioni scaturite dall'esempio della Terra Gemella trovano un riscontro anche nell'ipotesi (SENZ'ANIMA).

(SENZ'ANIMA), lo ribadiamo, è l'antecedente della tesi che Kim suppone sussista qualora accettassimo il monismo anomalo, rinunciando al riduzionismo senza volerci impegnare nel dualismo. In base a questa tesi, anche se alcune persone mancassero di proprietà mentali, finché tutte le loro proprietà fisiche e i loro ambienti fisici restano inalterati, si avrebbero gli stessi eventi fisici. Abbiamo visto come per Putnam tale tesi si basa su di un antecedente che egli battezza "l'ipotesi (SENZ'ANIMA)" secondo cui se qualche essere umano, pur avendo la nostra stessa struttura fisica, non ha proprietà mentali, questo non

246 Cfr. Putnam 1988, tr. it., p. 101.
247 Putnam 1999, pp. 86-87.

farebbe alcuna differenza sulle sue abitudini o sui suoi comportamenti.

Un fattore che Putnam identifica nell'affermare la mancanza di intelligibilità di (SENZ'ANIMA) è dato dal fatto che nessuno sa esattamente a cosa ci stiamo riferendo quando sosteniamo (SENZ'ANIMA), di cosa precisamente sono privi questi esseri umani. Con l'esempio della Terra Gemella Putnam ha mostrato come il significato di un termine è dato dalle relazioni che si hanno con l'ambiente e con gli altri parlanti, ma in questo caso pare che non sappiamo dire quale sia l'estensione del termine "stato mentale" (o "anima" o "coscienza"). Quello che ci manca, argomenta Putnam, è il concetto di coscienza, concetto che non acquisiamo dalla semplice consapevolezza di essere consci. Un concetto, come dimostrato dall'esperimento mentale della Terra Gemella, è qualcosa che acquisiamo dalle altre persone, ma è improbabile che la persona dalla quale ho appreso la parola (in questo caso "proprietà mentale" o "anima") non vuole dire niente usandola: «The language is a public language, after all»[248]; e ancora «il linguaggio è una forma di attività cooperativa, non un'attività essenzialmente individualistica»[249]. Pensare che quando sostengo (SENZ'ANIMA) intendo che "loro non hanno QUESTO", non ha un contenuto determinato finché il QUESTO non viene concettualizzato. Tuttavia non sapendo dare una definizione di anima, o di proprietà mentale, quali criteri usiamo per dire se qualcuno la possiede oppure no? Supporre che qualcuno sia "senz'anima" significa parlare fuori da tutti i criteri che abbiamo per coscienza, esperienza, ecc.

Inoltre Putnam afferma che quando incontra (SENZ'ANIMA) nel contesto di un'argomentazione come quella di Kim si trova in imbarazzo su come trattarlo, e questo poiché, argomenta l'autore, pare non si trovi un caso in cui si può utilizzare ciò che è descritto da (SENZ'ANIMA), proprio perché è assente un contesto, un parlato. In altre parole, il problema dell'antecedente di Kim non è nelle parole in quanto tali, ma nel fatto che per esse non è stato fornito un contesto che ci permetta di capire cosa si intende con esse; (SENZ'ANIMA) per Putnam non è assolutamente rilevante quando parliamo di causalità mentale: «the problem with Kim's antecedent is not in the words themselves; it is that we do not know what Kim means by them»[250].

Al riguardo Putnam riporta un esempio di Wittgenstein nel quale questi parla di una tribù priva di anima che un governo vuole schiavizzare proprio perché priva di anima, e

248 Ivi, p. 99.
249 Putnam 1988, tr. it., p. 47.
250 Putnam 1999, p. 91; cfr. anche Putnam 2012, cap. 28 *Wittgenstein: a Reappraisal*, pp. 486-488.

quindi utilizzabile senza scrupoli per qualsiasi proposito. Il riferimento di Wittgenstein qui era al popolo ebraico (la tribù senz'anima) e alla persecuzione che stava subendo da parte dei nazisti, tuttavia ciò che interessa a Putnam è sottolineare come in questo contesto si capisce benissimo cosa fanno le parole, che effetto intendono avere e che realmente hanno: «*in that context* I understand [soulless automata] all too well»[251]. La morale da trarre da questo esempio per Putnam è molto chiara: «"understanding" itself is context sensitive»[252]. Kim invece vuole trovare un modo per individuare le credenze senza appellarsi a fattori esterni, ma questo per Putnam non è possibile.

È qui che Putnam coglie l'occasione per farci notare come spesso le nostre forme di vita religiosa, e i discorsi intorno a esse nei quali è perfettamente coerente parlare di "anime disincarnate", "immortalità dell'anima", possano indurci a pensare che certe tesi filosofiche siano chiaramente intelligibili:

> what I am saying is that one reason we are tempted to accept [that there are soulless bodies that carry on just as we do] as an intelligible request is that having inherited a tradition in which it is (supposedly) intelligible to think of mental properties as "housed" in an immaterial soul, it naturally seems intelligible to ask what would happen if we "subtracted" that soul[253].

Tuttavia il pensiero religioso non supporta l'intelligibilità di un'ipotesi filosofica come (SENZ'ANIMA); il discorso filosofico può avere in parte ereditato dai miti religiosi la parvenza di intelligibilità di corpi senz'anima ma, sottolinea Putnam, i miti religiosi non dipendono per la loro funzione religiosa dal fatto che vi siano "mondi possibili nei quali le persone hanno proprietà fisiche ma non proprietà mentali".

3.5. Come definire e individuare una "credenza"

Putnam inoltre sostiene che il problema dell'interfaccia non riguarda solo la percezione, ma anche il pensiero (*conception*): per entrambi i concetti non avremo una definizione adeguata finché si accetta la caratterizzazione tradizionale della percezione. Se

[251] Putnam 1999, p. 91.
[252] *Ibid*.
[253] Ivi, p. 98.

esistesse l'interfaccia infatti, argomenta Putnam, non avremo una descrizione di come il linguaggio si aggancia al mondo: «The "how does the language hook on to the world" issue is, at bottom, a replay of the old "how does perception hook on to the world" issue. […] Is it any wonder that one can't see how thought and language hook on to the world if one never mentions perception?»[254]. In altre parole, data l'esistenza dell'interfaccia, non possiamo mai raggiungere direttamente gli oggetti e saremo dunque connessi a essi solo causalmente e non cognitivamente, ma quest'idea di percezione «makes it impossible to see how we can so much as *refer* to "external" things»[255].

Le credenze (i desideri, ecc.) sono, come si evince dall'esempio della Terra Gemella, individuate dal loro contenuto, e il contenuto di ogni frase dipende dal contesto nel quale viene espressa, ed è in genere individuato da condizioni che si trovano fuori dalla nostra testa:

> the content of beliefs […] is at least partly dependent on the determination of the *reference* in the particular context (in technical jargon, on the "extension") of the terms used in the sentence or in the expression of belief, and that reference depends on factors that are *external* to the speaker's body and brain[256].

Individuare una credenza equivale a individuare i contenuti delle parole e delle frasi che usiamo per formulare, comunicare e descrivere le credenze. Proprio come il significato di una credenza non può essere un ente solamente mentale, analogamente non può esserlo il contenuto di una credenza in quanto, come abbiamo visto, è possibile, a parità di stati cerebrali intrattenere credenze diverse in contesti ambientali differenti. Il contenuto di una credenza è il significato dell'enunciato usato per esprimerla: la tesi putnamiana sulle credenze segue quindi da quella originaria sul significato.

Dunque, quando diciamo "A crede che B", ciò che diciamo ha senso solo in relazione con il contesto nel quale lo diciamo. Per questo la credenza non può essere in assoluto uno stato interno, cioè uno stato che sopravviene a una proprietà specificabile in termini neurologici, fisici. Quindi, il possedere concetti non può essere uno stato funzionale, come invece sostiene Kim. Infatti quest'ultimo afferma che essere in un particolare stato psicologico è essere in uno stato con un ruolo particolare, ruolo che verrà realizzato da uno

254 Ivi, p. 12.
255 Ivi, p. 24.
256 Ivi, p. 119.

stato fisico. Dunque, dice Putnam, fondamentalmente Kim sostiene che ha senso pensare alle credenze (o agli stati mentali) come realizzate da stati fisici. Tuttavia, sottolinea l'autore, questo modo di vedere le cose non tiene in considerazione il modo in cui le credenze sono individuate, cioè in base al contesto, in base agli interessi del parlante, in base a cos'altro il parlante è disposto a dire su quella credenza, a quali altre credenze siamo disposti ad ascrivere al parlante. Caratterizzare gli stati mentali come fa Kim porta ad assumere che gli enunciati abbiano contenuti con valori di verità fissi, indipendentemente dal contesto. Rendersi conto di questo significò per Putnam il principio della fine del suo sostegno al funzionalismo: «But is the possession of concepts a functional state? (In my own thinking raising this question was the beginning of the end of my attachment to functionalism)»[257].

Putnam ritiene invece che «our psychological characteristics are, as a rule, individuated in ways that are context sensitive and extremely complex, involving external factors (the nature of the objects we perceive, think about and act on), social factors, and the projections we find it natural and unnatural to make»[258]. È la nostra pratica di interpretazione il criterio appropriato per identificare il contenuto di una credenza, desiderio, e così via. Il processo in base al quale formiamo le nostre credenze inizia, per Putnam, fuori dalla testa, ed è un esercizio di un intero sistema di funzioni, alcune formate dall'evoluzione, altre dai processi culturali, e tutti insieme connettono gli esseri umani agli oggetti e a ciò che accade nell'ambiente circostante:

> forming beliefs in accordance with our normal biological functions and our linguistic upbringing is not just uttering noises that are mere responses to sense impressions, although those sense impressions are a *part* of the causal chain that constitutes the normal formation of a particular perceptual belief on the basis of seeing something in one's visual field[259].

3.6. La falsa idea del mentale quale teatro interno

Come già detto in precedenza, Putnam ritiene che Kim caratterizzi le proprietà mentali

257 Ivi, p. 118.
258 Ivi, p. 132.
259 Putnam 2012, cap. 3 *Corresponding with reality*, p. 88.

come interne al nostro corpo, e afferma che è il miglior rappresentante di ciò che egli battezza cartesianismo e materialismo, riferendosi alla combinazione dell'idea cartesiana del mentale come teatro interno, ovvero all'assunzione che ci sia un regno autonomo di esperienze e fenomeni mentali, con l'aggiunta del materialismo.

Putnam, lo ribadiamo, è fermamente convinto che sia proprio l'errore nel concepire la percezione come teatro interno, come interfaccia cognitiva tra la mente e il mondo, a dare vita a numerosi problemi all'interno della filosofia della mente (e non solo), riprendendo le parole di Stalnaker, egli afferma che: «the exact content of the claim that a person believes so-and-so […] may depend on the context in which the claim is made in a way that is incompatible with the idea that what the claim means is that the person is in some one fixed "internal state"»[260].

La dottrina della percezione messa sotto accusa da Putnam è quella in base alla quale noi non percepiamo gli oggetti ma solo i dati di senso; il mondo agirebbe causalmente sulla mente, fornendogli input sensoriali, e la mente li processerebbe facendoli diventare degli output della forma appropriata. Inoltre, la concezione tradizionale della percezione postula l'argomento del *massimo divisore comune*: se ho la stessa esperienza in due occasioni, ma in una di queste non percepisco realmente ciò che mi sembra di percepire (per esempio sto sognando), allora ci deve essere qualcosa in comune nei due casi, e questo qualcosa in comune non può essere esterno, ma deve essere qualcosa interno che ho esperito. Questo qualcosa interno è chiamato rappresentazione, o apparenza, o dati di senso, ed è appunto il massimo divisore comune.

In base alla concezione tradizionale dunque, ciò a cui siamo cognitivamente relazionati nella percezione sono rappresentazioni. Queste rappresentazioni interne sembrano essere possedute dalle persone solo come effetti interni di cause esterne, e come riescano a fare ciò, dice Putnam, è qualcosa che è rimasto misterioso per secoli. Il problema, argomenta l'autore, è ritenere che ci sia una proprietà intrinseca, una proprietà che ha qualcosa di completamente indipendente dal linguaggio e dalla mente:

> it will do no good to search the brain for a "gran-mother neuron" – a point such that the messages from the visual cortex become *transformed* into sense data or qualia when they arrive *there*. Brains have speech centers, areas for different sorts

[260] Putnam 1999, pp. 117-18.

of memory, etc., but not consciousness centers[261].

L'inutile ricerca scientifica di oggetti scientifici chiamati "stati psicologici interni" è un esempio dell'errore di assumere che, se è naturale proiettare le stesse parole in due circostanze diverse, deve esserci un'entità che è presente in entrambe le circostanze.

L'unico modo che l'autore riconosce per uscire da tutto ciò è proprio abbandonare quest'idea della percezione, per sposare l'idea di una percezione immediata: «I believe that it is only by giving up this picture of perception as mediated by a set of "representation" in an inner theater that we will ever be able to escape from endless recycling of positions that do not work in the philosophy of mind»[262].

L'idea che Putnam ha della percezione è fortemente influenzata da pensatori quali James, Wittgenstein, Austin e McDowell, dai quali riprende l'idea che ciò che percepiamo non sono i dati di senso e che l'esperienza percettiva sia diretta.

La soluzione di Putnam (elaborata insieme a Hilla Jacobson) è il "*transazionalismo*", ossia l'idea che ciò che noi percepiamo «depends on a transaction between ourselves and the environment, […] the properties we perceive depend on our nature as well as the nature of the environment (This is why we call them "transactional" properties). […] I think of veridical experiences as *externalistically* characterized states of the organism, states that essentially involve the object perceived»[263].

La percezione quindi, secondo Putnam, avviene tramite un contatto cognitivo *diretto* tra noi e le proprietà degli oggetti esterni, e non tramite un qualche effetto o rappresentazione. Non bisogna pensare, argomenta Putnam, all'esperienza percettiva come a degli oggetti che stanno in un "teatro interno" che noi osserviamo con l'aiuto di capacità percettive interne al nostro organismo[264]: l'esperienza percettiva non è identica a stati del cervello.

Tra l'altro, la connessione con l'ambiente esterno per Putnam sussiste anche nei casi di percezione non diretta, come per esempio i sogni o le allucinazioni. Il modo giusto di trattare questi casi, sostiene Putnam, è quello di caratterizzarli in termini funzionali. Questo vuol dire che un'allucinazione (o un sogno) è uno stato nel quale al soggetto sembra di percepire qualcosa, ma non si tratta, evidentemente, di percezione diretta; percepire quel qualcosa non è

261 Ivi, p. 31.
262 Ivi, p. 102; cfr. anche Putnam 2012, cap. 34 *Aristotle's Mind and the Contemporary Mind*, p. 590.
263 Ivi, cap. 36 *How to Be a Sophisticated "Naïve Realist"*, pp. 636-638; cfr. anche Putnam 1999, p. 159.
264 Cfr. Putnam 2012,cap. 36 *How to Be a Sophisticated "Naïve Realist"*, p. 627.

uno stato che riguarda il mondo in maniera diretta, ma è la sua caratterizzazione funzionale che si riferisce allo stato che riguarda il mondo in maniera diretta. Per questo anche nei casi di percezione non diretta è presente un contenuto: «it is part of this "functionalist" view that ("philosophical") hallucination do have content»[265].

Gli oggetti di un'esperienza non veridica dobbiamo trattarli al pari di entità immaginarie, «the hallucinated [object] has the ontological status of a *fictional entity*»[266]. Quando abbiamo a che fare con personaggi immaginari, come per esempio alcuni personaggi che ci offre la letteratura, quei personaggi, nonostante non siano reali, ci *sembrano reali*. Allo stesso modo, nel caso delle allucinazioni, l'oggetto che non è reale sembra tuttavia reale per il soggetto che ha l'allucinazione. Le allucinazioni non coinvolgono l'oggetto percepito, esse piuttosto rappresentano un oggetto come se fosse percepito, e questo perché le allucinazioni sono stati difettosi, sono malfunzionamenti. Al contrario, l'esperienza veridica coinvolge l'oggetto, non lo rappresenta semplicemente. Questo, secondo Putnam, è il modo migliore per rendere conto dell'esperienza diretta e non diretta, senza far uso di mediatori come i dati di senso.

3.7. La sopravvenienza

Questo discorso per Putnam è rilevante poiché egli ritiene che la concezione della sopravvenienza psicofisica di Kim presupponga il tipo di percezione appena criticata, ovvero quello di stampo cartesiano. Per poter sostenere la sopravvenienza forte senza negare l'esternismo semantico Kim, argomenta Putnam, è stato costretto a dividere gli stati psicologici non interni (come per esempio "credere che ci sono chiese a Vienna") in uno stato interno e in una parte di relazioni esterne storiche e cognitive, concludendo che la sopravvenienza vale solo per la parte interna dello stato psicologico non interno complessivo. Il problema, sottolinea Putnam, è se capiamo cosa viene detto quando affermiamo, per esempio "credo che ci sono chiese a Vienna" è «*an internal psychological state with the same causal-explanatory role as the noninternal state* of knowing that there are churches in Vienna»[267]. L'argomentazione di Kim, per Putnam, ruota attorno a concetti del tutto

265 Ivi, p. 637.
266 *Ibid*.
267 Putnam 1999, p. 113.

inintelligibili, quali proprietà psicologica interna, stati fenomenologici interni e stati funzionali.

Proprio perché non c'è un'interfaccia cognitiva tra gli esseri umani e il mondo, la nostra relazione con il mondo è diretta, e questo porta al fatto che non ci sia uno stato mentale interno o, detto diversamente, il contenuto stretto, ovvero quel contenuto che sopravviene su ciò che accade all'interno della testa dell'individuo. Supporre che ci sia qualcosa come il contenuto stretto è supporre che vi sia un'interfaccia cognitiva tra noi e il mondo, e questa abbiamo visto essere un'idea che Putnam non accetta più.

Quando negli anni '70 [268] Putnam inizia a parlare di contenuto ampio e contenuto stretto, egli riteneva che la nozione di mente fosse ambigua e che in un certo senso di stato mentale (nel senso stretto) i nostri stati mentali fossero interamente nelle nostre teste; e in un altro senso (nel senso ampio) che includeva stati quali conoscere il significato di una parola, i nostri stati mentali fossero individuati dalle nostre relazioni con l'ambiente e gli altri parlanti, e non solo da ciò che c'è nella testa. Ora tuttavia Putnam crede che tutto ciò conceda troppo all'idea che la mente può essere concepita come un teatro interno, situato all'interno della testa[269].

Putnam individua tre problemi fondamentali quando si parla di contenuto stretto: 1. il significato delle parole è in parte individuato da fattori esterni; 2. il significato è sensibile al contesto; 3. il modo olistico con il quale determiniamo quando due enunciati hanno lo stesso significato e quando invece lo hanno diverso, ovvero il fatto che non si può determinare il significato di un enunciato considerandolo in maniera isolata, in quanto il significato di un enunciato dipende dalle connessioni che tale enunciato intrattiene con il resto del linguaggio. Questi tre elementi rendono per Putnam impossibile parlare di contenuto stretto o di stato mentale interno.

Tuttavia è importante notare come Putnam non rifiuta completamente la nozione di sopravvenienza, accettando un'idea di sopravvenienza globale che non si riferisce alla correlazione tra stati mentali e stati fisici:

> I do think that all of our capacities, including "agential" ones (a category which […] includes our perceptual capacities), supervene on the states of the physical

[268] Cfr. Putnam 1975, cap. 12 *The Meaning of 'Meaning'*, p. 220.
[269] Cfr. Pessin e Goldberg 1996, p. xviii.

universe[270].

Pertanto non è uno stato mentale interno a sopravvenire su uno stato fisico del cervello: in Putnam la relazione di sopravvenienza è tra le nostre capacità di esseri umani e l'ambiente fisico che ci circonda. Qui ritroviamo sia la concezione della mente di Putnam sia il suo esternismo: la mente infatti non viene intesa come interna al nostro cervello, non come oggetto che debba essere sottoposto a indagine scientifica, ma come un sistema strutturato di abilità che sopravvengono sui fattori esterni all'organismo, dunque sopravvengono sull'ambiente naturale esterno al soggetto, ambiente che abbiamo visto giocare un ruolo centrale nel determinare il contenuto dei nostri stati mentali.

Come ci fa notare lo stesso Putnam, questa caratterizzazione della sopravvenienza è compatibile anche con la sua idea di percezione diretta transazionale: «perception is *transactional* [the term is Dewey's]. And that is compatible with supervenience, because transactions between the eye and the things 'out there' that we see are also material. Seeing a tree is supervenient on material processes»[271]. La percezione, argomenta Putnam, è sopravveniente ai processi materiali, ma questi non sono da intendere come processi nel cervello, altrimenti torniamo all'idea che la mente sia una cosa che sta dentro le nostre teste e che i processi cognitivi siano confinati nel cervello.

Putnam quindi respinge l'idea che la tesi della correlazione psicofisica abbia un senso, poiché presuppone un'immagine sbagliata, ossia quella secondo cui le nostre caratteristiche psicologiche sono stati interni correlati con ciò che accade dentro ai nostri corpi. Tutto ciò cui di solito si fa appello per supportare l'idea che le credenze siano stati interni, afferma Putnam, è pura fantascienza.

3.8. Il realismo diretto (o naturale)

Non ci soffermeremo qui a parlare nel dettaglio del realismo sostenuto attualmente da Putnam; ai fini del nostro lavoro è importante però sapere che questi ora difende un tipo di realismo che va sotto il nome di realismo diretto o naturale. Seguendo Maria Baghramian[272]

[270] Putnam 2008, p. 30.
[271] Putnam 2002, p. 125.
[272] Cfr. Baghramian 2008, p. 30.

possiamo così riassumere alcune caratteristiche di questo realismo: gli oggetti della percezione sono cose esterne, o aspetti di una realtà esterna, e le nostre capacità cognitive possono raggiungere direttamente gli oggetti stessi, dunque le cose esterne possono essere esperite direttamente dal soggetto, senza passare per i dati di senso; il contenuto delle nostre parole dipende dalle nostre relazioni con il mondo, dal modo in cui siamo inseriti in una cultura e nell'ambiente fisico; gli esseri umani utilizzano il linguaggio per vari scopi, incluso quello di descrivere la realtà, tuttavia le descrizioni del mondo non sono mai una semplice copia di questo. Il mondo è quello che è indipendentemente dagli interessi di chi lo descrive: il nostro mondo e la nostra vita «are constrained by a reality not of our own invention»[273]. Tuttavia i nostri interessi sono fondamentali nel determinare ciò che crediamo, pensiamo, desideriamo e così via.

Per Putnam il realismo naturale ci fa capire «the *needlessness* and the *unintelligibility* of a picture that imposes an interface between ourselves and the world»[274]. Inoltre per Putnam il realismo diretto

> is best thought of not as a *theory* of perception but as a denial of the necessity for and the explanatory value of positing "internal representations" in thought and perception[275].

Come ci fa notare la Baghramian, il fulcro del realismo naturale di Putnam è il suo tentativo «to do justice to our sense that knowledge claims are responsible to reality without recoiling into metaphysical fantasy»[276].

[273] Putnam 1999, p. 9.
[274] Cfr. ivi, p. 41.
[275] Ivi, p. 101.
[276] Cfr. Baghramian 2008, p. 29; cfr. anche Putnam 1999, p. 4.

CAPITOLO 4

Stephen L. White: la a-sopravvenienza

4.1. La critica di Stephen White

Interessante è la critica mossa da Stephen White nei confronti di Putnam. White si trova d'accordo con Putnam sul fatto che non ci sia un'interfaccia tra gli oggetti del mondo esterno e le nostre menti, tuttavia l'argomentazione utilizzata da Putnam per dimostrare l'assenza di tale contenuto è per White incompleta, e quindi inefficace.

Come abbiamo illustrato nei paragrafi precedenti, Putnam ritiene di aver dimostrato attraverso l'esperimento di Terra Gemella che il contenuto dei nostri pensieri è fissato in modo "ampio", ovvero dall'ambiente e dalla società linguistica nella quale viviamo, e non in modo "stretto", cioè da ciò che accade nelle nostre teste.

In base a questa caratterizzazione, il rapporto tra mentale e fisico è per Putnam quello della sopravvenienza globale, riferendosi alla sopravvenienza delle capacità degli esseri umani su fattori esterni all'organismo, sugli stati fisici dell'universo (e non del soggetto). L'idea del contenuto stretto invece porterebbe con sé l'idea di sopravvenienza che possiamo definire locale, in base alla quale gli stati mentali interni sopravvengono sugli stati fisici del cervello

White, riportando proprio l'esempio di Terra Gemella, sottolinea che: «granting that our beliefs have broad content that does not supervene on what is inside the subject's head does not preclude their having a kind of content that does»[277]. L'esistenza del contenuto

[277] White 2008, p. 12.

ampio non esclude la possibilità «of our factoring such content into a narrow content plus an external causal chain. […] The factors that determine the association between a term and, say, a natural kind for a particular context of acquisition were like the reference determining factors that figure in typical causal theories. And Putnam provides no argument that narrow content should *not* presuppose broad content in this way»[278].

Di conseguenza, finché Putnam non dimostra in maniera convincente che non c'è qualcosa come un contenuto stretto che presuppone quello ampio, non avrà dimostrato neanche la plausibilità del realismo diretto. Infatti il realismo diretto, come si è visto, si basa per Putnam sull'assenza di un'interfaccia tra parlante e oggetto, interfaccia che però continuerebbe a sussistere se il contenuto ampio fosse determinato in virtù di un legame causale, legame che costituirebbe appunto un'interfaccia tra la mente e il mondo.

Nella visione di Putnam, se ci fosse una qualche interfaccia non capiremo come il linguaggio si aggancia al mondo dato che, posto che a un capo dell'interfaccia vi è la mente umana, nulla assicura che all'altro lato vi sia il mondo (o meglio, un suo aspetto). Da ciò segue che anche la possibilità di avere un linguaggio significante viene messa a repentaglio. Per questo Putnam pensa che il rapporto percettivo ed epistemico con il mondo non può non essere diretto. Parallelamente Putnam ritiene che a garantire riferimento e significato a buona parte dei nostri termini vi sia una quantità di catene causali – una per ciascun termine - , dove ogni catena collega gli usi che i parlanti fanno di quel termine con gli oggetti a cui il termine si riferisce. Dunque le catene sono fondamentali per avere un linguaggio significante.

Tuttavia per White l'esistenza della catena causale (che per Putnam conferisce sostanza all'idea di contenuto ampio) non è sufficiente a escludere l'esistenza del contenuto stretto. In più, la catena causale costituirebbe proprio quell'interfaccia che Putnam cercava di escludere, in quanto potrebbe essere la "barriera" che media tra soggetto e oggetto. Secondo White quindi non è sufficiente introdurre la nozione di catena causale per eliminare completamente la "minaccia" della presenza del contenuto stretto. Infatti, sebbene l'idea della catena causale aiuti a fornire una spiegazione del modo in cui il linguaggio si aggancia al mondo, il problema di questa, sottolinea White, è che non è accessibile cognitivamente al soggetto, dunque non rende giustizia alla sua razionalità. Il parlante difatti non percepisce la catena causale, non ha sostanzialmente nessuna esperienza diretta con la catena causale, non c'è sostanzialmente nessuna prova accessibile al soggetto dell'esistenza della catena causale.

[278] Ivi, pp. 12-13.

Come ci spiega l'autore, per la teoria causale il riferimento è indipendente dalla descrizione disponibile al soggetto; se un soggetto crede due cose diverse sullo stesso oggetto, il problema è che in base alla teoria causale queste due credenze sono sullo stesso oggetto e dicono cose contradditorie, facendo sembrare il soggetto irrazionale. Tuttavia, argomenta White, non possiamo contrastare questa implicazione facendo ricorso alle diverse descrizioni che il soggetto associa ai due termini, fare questo minerebbe la pretesa della teoria causale che il riferimento è indipendente dalle descrizioni disponibili al soggetto. Né possiamo fare appello alla catena causale che collega, per esempio, Venere con "stella del mattino" e "stella della sera", poiché la catena causale non è disponibile al soggetto. Le catene causali quindi non razionalizzano né giustificano il comportamento del soggetto, proprio perché la catena causale non è accessibile cognitivamente al soggetto.

Per White ci vuole qualcosa di diverso per dimostrare che non è il contenuto stretto ma quello ampio a determinare il contenuto delle nostre credenze, qualcosa che non si basi su alcuna interfaccia. Ci vuole un argomento trascendentale, un argomento cioè che sottolinei le condizioni di possibilità del riferimento.

Ciò che White propone è un modo per riconciliare il realismo diretto con il vincolo di Frege. Con tale espressione egli intende il fatto che se un soggetto ha due credenze riguardo a un oggetto, credenze che logicamente non possono essere vere dello stesso oggetto (vedi l'esempio precedente di Venere), allora ci devono essere due diversi modi di presentazione di quell'oggetto in base ai quali è possibile avere due credenze diverse sullo stesso oggetto. Spieghiamo meglio di cosa si tratta.

Frege riteneva che uno stesso oggetto potesse essere descritto in più modi e la diversità del significato dei termini che si riferiscono a uno stesso oggetto è data dai diversi modi in cui presentano l'oggetto. Per riprendere il solito esempio, enunciati come "Fosforo è la stella del mattino" e "Espero è la stella della sera", si riferiscono allo stesso oggetto (Venere) ma hanno significati distinti, tuttavia questa diversità nel significato è data dalla diversa esperienza (i modi di presentazione) che un soggetto ha nei confronti di Venere.

In realtà, per White l'utilità del ricorso al vincolo di Frege è più evidente negli enunciati dimostrativi, gli enunciati *de re*, ovvero quelli che dipendono dall'oggetto. Per maggiore chiarezza riportiamo un esempio di White, che a sua volta lo riprende da G. Evans: supponiamo che qualcuno guardi fuori da un oblò verso destra, vede la prua di una nave e dice "quella nave è di fabbricazione giapponese". Poi si volta sulla sinistra vedendo attraverso

un altro oblò la poppa e afferma "quella nave invece no". Senza rendersene conto ha indicato due volte la stessa nave e detto cose che sono logicamente incompatibili su di essa. La diversità nella percezione dello stesso oggetto (la nave) è data dai diversi modi di presentazione, cioè dalle diverse esperienze coscienti – e non dalla catena causale – che il soggetto ha verso l'oggetto. Ora, come si relaziona tutto questo con il realismo diretto?

Come detto, White non crede che quello della catena causale sia un buon argomento per eliminare l'idea di un'interfaccia tra i nostri pensieri e gli oggetti del mondo esterno. La soluzione per White è da ricercarsi in quelli che egli definisce «packages of basic action possibilities»[279]. Un'azione basilare è un'azione che non implica modi di presentazione neutrali rispetto al mondo esterno, nel senso che per avere esperienze coscienti c'è bisogno degli oggetti del mondo verso i quali compiere le nostre azioni basilari. Le azioni basilari sono dunque direttamente collegate all'oggetto, sono tutte quelle azioni che, come detto, un soggetto non potrebbe compiere se non ci fossero gli oggetti del mondo esterno verso i quali indirizzarle, sono quindi quelle azioni che si compiono direttamente verso l'oggetto o, detto diversamente, sono tutte quelle azioni per le quali il soggetto che le compie non ha bisogno di fare altro per essere collegato all'oggetto; gli oggetti offrono intrinsecamente possibilità di azioni e manipolazioni. Riallacciandoci all'esempio precedente della nave dove, lo ricordiamo, un soggetto aveva inconsapevolmente due esperienze diverse e contraddittorie sulla stessa nave, secondo White questo esempio mette in luce le diverse possibilità di azioni basilari, a seconda di dove si guarda la nave, se a prua o a poppa. In entrambi i casi tuttavia, sottolinea White, il soggetto può indicare la nave, tracciarne con il dito il perimetro nell'aria, avvicinarsi per avere una visione migliore, e fare un'indefinita quantità di altre cose, tutte definibili come azioni basilari poiché «one simply performs the actions directly on "that ship"»[280].

Le nostre esperienze coscienti sono dunque vincolate all'esistenza degli oggetti nel mondo esterno, oggetti che ci offrono possibilità di azioni basilari. I due modi di presentazione della nave sono, afferma White «a matter of know how, and this is crucial to the claim they are object involving»[281].

L'analogia con le azioni basilari aiuta quindi a spiegare come sia possibile avere un linguaggio significante. Infatti, le azioni basilari danno contenuto semantico all'oggetto,

279 Ivi, p. 20.
280 Ivi, p. 21; cfr. anche White 2007b, p. 4.
281 White 2008, p. 21.

proprio perché sono indirizzate direttamente all'oggetto. E il fatto che nel mondo ci siano degli oggetti è ciò che fornisce le condizioni di possibilità dell'esistenza del linguaggio, se non ci fossero gli oggetti non vi sarebbe neanche il linguaggio:

> *de re* (or object-involving, or Russellian) thought is necessary to our having a meaningful language and […] such object–involving thought is irreducible to thought that is non-object-involving. (To say that some thought or intentional content is object involving is to say that if the object or objects in question do not exist there is no complete thought; there is nothing that could be evaluated for truth, accuracy, veridicality, or the like)[282].

Queste riflessioni di White sono influenzate dal pensiero di J. J. Gibson, dal quale White riprende l'idea che ciò che ci viene dato dall'esperienza non è un mondo di oggetti e strutture neutrali rispetto all'azione, ma un mondo di opportunità di azioni fornite dagli oggetti, «a world of affordances – of shelters and hiding places, doorways, stairs, bridges, and scape routes»[283].

Le cose che vediamo direttamente sono dunque per White le cose che possiamo manipolare attraverso le nostre azioni basilari, quindi la percezione diretta è possibile grazie a questo, e non alla catena causale. Dobbiamo avere qualche credenza percettiva che riguarda un oggetto per avere un linguaggio significante, per White «our access to the world, is less epistemological than practical – less a matter of knowledge than of know how»[284].

I pensieri che riguardano gli oggetti sono necessari per avere un linguaggio significante: «we must have some object-involving perceptual beliefs in order to have a meaningful language»[285]; e il fatto che riguardano gli oggetti consente che la relazione dei nostri pensieri con il mondo sia diretta.

4.1.1. Cosa non va con la sopravvenienza

Consideriamo ora la questione della sopravvenienza in relazione all'argomento

[282] Ivi, p. 15.
[283] Ivi, p. 22.
[284] Ivi, p. 27.
[285] Ivi, p. 23.

trascendentale di White. Egli esclude la possibilità della sopravvenienza degli stati mentali sugli stati fisici del cervello poiché porterebbe con sé l'idea dell'interfaccia tra l'interno e l'esterno che White non accetta; inoltre esclude anche la possibilità della non sopravvenienza degli stati mentali sugli stati fisici perché «threatens to make facts about what it is like to have one's present experiences mysterious»[286].

White sostiene molto esplicitamente che il problema della sopravvenienza «cannot arise, since there is no common ground from which to ask it»[287]; secondo l'autore il piano mentale e il piano fisico sono *incommensurabili* nel senso di Kuhn. Quest'ultimo, per farla breve, riteneva che due teorie sono incommensurabili quando «there is no language, neutral or otherwise, into which both theories, conceived as sets of sentences, can be translated without residue or loss»[288]. Dunque, dire che due teorie sono incommensurabili significa per Kuhn ritenere che non c'è alcun piano su cui esse possano essere, per così dire, collocate e poste a confronto, perché ogni confronto del genere presuppone qualcosa in comune, per esempio un linguaggio, oppure un insieme di significati, oppure un insieme di concetti, qualcosa insomma che possa permettere di esprimere ciò che ciascuna teoria afferma, in una maniera che risulti comprensibile ai sostenitori dell'una come dell'altra teoria. Quando questo qualcosa di comune tra esse non esiste, allora le due teorie non possono essere confrontate tra loro: sono appunto incommensurabili[289].

Relazionato a un eventuale rapporto di sopravvenienza, l'incommensurabilità di cui parla White si riferisce al fatto che tra mentale e fisico non c'è una base comune: mentale e fisico sono dimensioni così radicalmente differenti da generare ciascuna una propria *base percettiva*, ossia entrambe le prospettive ci danno una visione del mondo completamente diversa; ciò che percepiamo ci apparirà in modo diverso se consideriamo la realtà come solamente fisica o come completamente mentale. Non solo: secondo White, qualora considerate nella loro globalità, ciascuna delle due basi minaccia l'altra, nel senso che si escludono a vicenda: «each perspectives generates its own perceptual base. […] the relation between the two perspectives is dialectical – each grows naturally out of the other, and each, if carried to its logical extreme, undermines (or threatens to undermine) the other»[290]. Sembra

286 Ivi, p. 26.
287 Ivi, p. 27.
288 Kuhn 1982, p.670.
289 *Ibid*.
290 White 2008, p. 26.

dunque che una visione in termini fisicalistici del mondo non lasci spazio a tutto ciò che riguarda la sfera del mentale, viceversa una visione in termini mentali non lascio spazio alcuno al fisico.

Per quanto riguarda la percezione, White specifica che secondo lui la percezione è carica di teoria, e dunque le cose ci possono apparire diverse a seconda delle teorie che il soggetto ha interiorizzato e se viste da una prospettiva soggettiva o oggettiva. Ora il fatto è che, come dimostra la nostra esperienza, non abbiamo una e una sola prospettiva nella sua completezza, quella agentiva o quella oggettiva, ma una combinazione delle due. Il fisico (o l'oggettivo come preferisce definirlo White) nasce dal mentale (o dall'agentivo), poiché per capirci meglio in quanto agenti ci oggettiviamo rendendoci appunto oggetto di indagine (delle scienze umane). Dunque, ci concepiamo come oggetti nel tentativo di acquisire una comprensione migliore di noi come agenti. Viceversa l'agentivo cresce dall'oggettivo, ovvero dal nostro bisogno di rendere giustizia agli aspetti mentali che caratterizzano la nostra vita e che non sono riducibili (o riconducibili) alla sfera fisica. Questo intreccio tra le due prospettive testimonia il fatto che le cose non ci possono essere date sotto un aspetto totalmente agentivo e totalmente oggettivo contemporaneamente, in quanto le due dimensioni sono così diverse che non hanno nulla a che vedere l'una con l'altra. Se nel descrivere la realtà consideriamo, per esempio, solo l'aspetto oggettivo lasceremo fuori tutto ciò che riguarda il mentale; diversamente, se consideriamo la realtà come tutta mentale, allora non lasceremo spazio alla dimensione fisica.

Quindi, sembra che una prospettiva agentiva definita sia impossibile da riconciliare con l'assunzione che l'oggettività scientifica ci dà la prospettiva ultima del mondo. Se riteniamo che l'oggettività scientifica ci dà tutte le risposte a tutte le domande sui fatti oggettivi allora, argomenta White, sembra non esserci posto per l'agentivo, il soggetto viene completamente escluso. L'oggettivo, pertanto, precluderebbe l'agentivo, escludendo la possibilità di concetti agentivi dotati di significato; se vediamo il fisico come completo, come esaustivo, allora l'azione del soggetto verrà minacciata, ovvero sembra che non ci sia spazio per la libertà di azione. Tuttavia, rendere giustizia all'azione del soggetto comprometterà o l'oggettività in terza persona dell'oggettivo o l'assunzione che tale oggettivo fornisce una prospettiva ultima sul mondo.

L'incommensurabilità tra agentivo e oggettivo porta dunque secondo White a sostenere la *a-sopravvenienza*, ovvero il fatto che, a causa della mancanza di una base comune tra i due,

non ha *senso* chiedersi se ci possano essere mondi in cui i fatti fisici sono gli stessi, ma con diversi fatti mentali: «the position, then, is not supervenience or non supervenience, but asupervenience»[291]. Il mentale è irrilevante per il fisico poiché appartengono a prospettive incommensurabili. L'argomento trascendentale mostrato in precedenza e la tesi della a-sopravvenienza sono per White fortemente legati: «siccome abbiamo un argomento trascendentale per cui dobbiamo applicare concetti agentivi per rendere sensata la possibilità di avere un linguaggio dotato di significato, possiamo avere un argomento trascendentale contro il nostro ritenere totalizzante lo schema concettuale oggettivo»[292].

Proprio perché l'argomento trascendentale e la tesi della a-sopravvenienza vanno di pari passo, secondo White la sopravvenienza globale sostenuta da Putnam è altrettanto inaccettabile quanto la sopravvenienza locale e la non sopravvenienza. Infatti, argomenta White, è compatibile con la sopravvenienza globale il fatto che il mentale aleggi libero da ogni fatto fisico *rilevante* poiché è sufficiente per la sopravvenienza globale che se due mondi differiscono per un aspetto mentale, allora ci deve essere una differenza fisica fra i mondi, anche se questa differenza è tanto irrilevante quanto una differenza tra due molecole – una in un mondo, l'altra nell'altro[293].

Il problema che White identifica nella sopravvenienza è che quella della sopravvenienza è una relazione che non riesce a «spiegare o rendere intelligibile *come* le proprietà sopravvenienti derivano dalle proprietà di base, né è sufficiente per garantire la "innocenza ontologica" delle proprietà sopravvenienti»[294]. Ciò che la sopravvenienza non garantirebbe quindi è l'intuizione che ci sia una "*rilevanza positiva*" del fisico sul mentale. Se sembra assurdo, argomenta White, che ci possa essere una distinzione mentale senza alcuna differenza fisica, sembra altrettanto assurdo che ci possa essere una distinzione senza una differenza *rilevante*. Ma nella misura in cui questa intuizione ha la sua fonte nel reame del mentale, conclude White, la sopravvenienza fallisce[295].

La posizione che si viene delineando ci pare quella di un dualismo: se mentale e fisico sono incommensurabili allora siamo di fronte a due dimensioni diverse. Nel primo capitolo abbiamo descritto i diversi tipi di dualismo presenti nel panorama filosofico attuale,

---

291 Ivi, p. 27.
292 White 2010, p. 195.
293 Cfr. ivi, p. 196.
294 *Ibid*.
295 Cfr. ivi, pp. 197-198.

illustrando la macro divisione tra dualismo ontologico e dualismo epistemologico, e l'ulteriore suddivisione del dualismo ontologico in dualismo delle sostanze e delle proprietà. Ci pare che quello descritto da White si possa avvicinare al dualismo ontologico *à la* Chalmers, cioè al dualismo delle proprietà. White infatti non pare alludere a due sostanze differenti quando definisce mentale e fisico incommensurabili: «i più chiari esempi in cui una classe di entità aleggia libera da un'altra – per esempio menti e cervelli cartesiani – sono quelli in cui entrambe le classi appartengono alla stessa categoria (sostanze)»[296]; piuttosto pare riferirsi a caratteristiche distinte dei due elementi. Lo ricordiamo, in base al dualismo delle proprietà la sostanza può essere la stessa, ma il mentale e il fisico sono aspetti separati e irriducibili l'uno all'altro. Quanto sostenuto da White sembrerebbe rientrare in questo tipo di dualismo: egli infatti, seppur non parlando esplicitamente di proprietà, sottolinea come mentale e fisico siano due dimensioni talmente diverse da non avere niente in comune, e quindi definibili come due aspetti separati e irriducibili l'uno all'altro. White parrebbe ipotizzare due dimensioni di analisi degli esseri umani, una oggettiva e l'altra agentiva che, come abbiamo appena illustrato, non hanno niente a che vedere tra di loro proprio perché sono incommensurabili.

Anche la tesi della a-sopravvenienza sembra andare a favore di una visione dualista della realtà: è proprio a partire dall'incommensurabilità tra le due sfere, dalla mancanza di possibilità di trovare un rapporto, una relazione tra mentale e fisico che White avanza la sua tesi. Una tesi che nel cercare di rendere giustizia all'agentivo lo vede come completamente diverso, imparagonabile e soprattutto irrilevante per l'oggettivo, lasciando spazio all'ipotesi dell'esistenza di due dimensioni distinte.

Se White avesse ragione il suo argomento potrebbe segnare un punto di svolta nel discorso sulla sopravvenienza e sul correlativo problema mente-corpo: il fatto che mentale e fisico siano incommensurabili, il fatto che ciascuna dimensione sia ininfluente per l'altra, potrebbe spiegare perché sembra così complicato riuscire a capire quale relazione ci sia tra i due ambiti. La risposta di White sarebbe presumibilmente che non c'è nessuna relazione poiché non sussiste proprio il problema di relazionare due sfere incommensurabili.

4.1.2. La risposta di Putnam

[296] Ivi, p. 200.

Putnam ha avuto modo di fornire una breve risposta alle critiche mossegli da Stephen White, una risposta che si focalizza sulla sopravvenienza. Egli infatti continua a difendere la sopravvenienza globale, sostenendo che «in the case of a presentation of a public property or a public object, I do not see Frege's constraint as driving us to reject global supervenience at all»[297].

Immaginiamo, argomenta Putnam, che ci sia un pianeta, supponiamo di chiamarlo Terra Invertita, dove i colori sono invertiti, per esempio il cielo è giallo, l'erba è rossa e così via. Tuttavia, gli abitanti hanno sviluppato, con l'evoluzione, un sistema visivo che funziona come una lente che inverte i colori, quindi il cielo, per esempio, anche se è giallo verrà percepito azzurro, l'erba anche se rossa verrà percepita verde e così via. Poniamo ora che un terrestre venga portato su Terra Invertita inconsapevolmente e senza conoscere le differenze sui colori; e supponiamo che, sempre a sua insaputa, gli vengano messe delle lenti a contatto che invertono i colori, in modo che egli possa percepire i colori esattamente come sulla Terra. L'abitante di Terra Invertita e il terrestre guardando il cielo affermeranno "Il cielo è blu", ma dato che il terrestre sta parlando l'italiano che si parla sulla Terra, mentre l'abitante di Terra Invertita sta parlando l'italiano di Terra Invertita, il terrestre si sta sbagliando, infatti il cielo su Terra Invertita è giallo, non blu. Quindi quando il terrestre, concorde con l'abitante di Terra Invertita, afferma che il cielo è blu, egli dà una descrizione del cielo che è incompatibile con la sua, in quanto l'abitante di Terra Invertita vuol dire che il cielo è giallo (nell'italiano di Terra), o perlomeno giallo è il colore a cui la parola "blu" si riferisce nell'italiano di Terra Invertita. Ora, prosegue Putnam, se applichiamo il vincolo di Frege, dovrebbe esserci una differenza rilevante nel modo in cui si presenta il cielo al terrestre e all'abitante di Terra Invertita, differenza che ci deve spiegare il fatto che il terrestre abbia accettato due enunciati (il suo e quello dell'abitante di Terra Invertita) incompatibili riguardanti il colore del cielo. Questa differenza, osserva Putnam, sta nel fatto che il terrestre sta vedendo il cielo attraverso lenti a contatto che invertono i colori, ma questa non è una differenza negli stati cerebrali del terrestre e del nativo di Terra Invertita. In base al vincolo di Frege, dunque, potremo dire che «the modes of presentation of the color of the sky is supervenient on the brain-state *plus* something external to the brain: the presence of contact lenses»[298] (per lo meno nel caso del terrestre).

[297] Putnam 2008, p. 32.
[298] Ivi, p. 31.

La coscienza di oggetti e proprietà pubblici, come la coscienza dell'abitante di Terra Invertita e del terrestre del colore dei loro rispettivi cieli quando si trovano nei loro pianeti, è per Putnam mediata da capacità oggettuali, capacità che sono stati funzionali, nel senso di funzionale sostenuto dall'ultimo Putnam, e queste capacità funzionali «have long arms; they reach out to the world»[299].

Egli tuttavia non replica alle accuse di White sul fatto che la presenza della catena causale del riferimento possa essere vista come quell'interfaccia tra mente e mondo che Putnam fortemente ha negato esistere, lasciando così la questione aperta.

[299] Ivi, p. 32.

Conclusioni

Abbiamo iniziato la discussione ponendoci il problema della relazione tra mente e corpo, ovvero chiedendoci quale sia la relazione ontologica tra mentale e fisico o, detto diversamente, come sia possibile che un mondo che sembra fatto essenzialmente di materia possa dar vita a stati mentali come desideri, credenze, speranze, che sembrano non avere niente di fisico.

Si sono illustrate alcune delle soluzioni che nei secoli ha cercato di dare la filosofia, partendo dal dualismo cartesiano, passando per le prime posizioni riduzioniste, fino ad approdare alle posizioni più recenti, come il funzionalismo, influenzate sempre più dalle relazioni tra filosofia e neuroscienze.

Da questa analisi è emerso che un ruolo importante nel rendere conto della relazione tra mentale e fisico, da Davidson in poi, ha avuto la nozione di *sopravvenienza*. Egli introdusse questo concetto per dare una spiegazione non riduzionista di come la sfera psicologica interagisce con la sfera fisica. Nella visione di Davidson, le proprietà mentali non si riducono alle proprietà fisiche, anzi le proprietà mentali di un evento dipendono dalle proprietà fisiche di quell'evento poiché le proprietà mentali sopravvengono sulle proprietà fisiche, nel senso che due eventi non possono essere uguali in ogni aspetto fisico e differire in qualche aspetto mentale. Abbiamo anche sottolineato che la sopravvenienza proposta da Davidson è stata definita "debole", perché la relazione tra le caratteristiche mentali sopravvenienti sulle caratteristiche fisiche non vale, dal punto di vista ontologico, in tutti i mondi possibili.

Si è mostrato come la sopravvenienza intesa in questi termini abbia creato notevoli questioni. Particolarmente convincente ci è parsa l'argomentazione di Kim, il quale ritiene che tale nozione, così intesa, anziché salvaguardare il ruolo causale del mentale relega il mentale a mero epifenomeno, o addirittura lo elimina del tutto. Questo perché, nel sopravvenire al fisico, sembra che il mentale abbia un ruolo solo in quanto esemplificato dal fisico, e non in quanto mentale. Per Kim, nel tentativo di sostenere una posizione anti-riduzionista e senza voler ricadere nel dualismo, Davidson finisce per non dare alcun aspetto rilevante alle proprietà mentali.

Ciò che è emerso è che nella visione di Kim l'unica via che si ha per salvaguardare il mentale è quella del riduzionismo, un riduzionismo funzionalizzato che ricorre comunque alla nozione di sopravvenienza per spiegare la relazione tra mentale e fisico. Egli parla infatti di sopravvenienza "forte" o causazione sopravveniente, in base alla quale una proprietà mentale sopravviene alla proprietà fisica corrispondente perché quella proprietà fisica *la realizza*. Kim si riferisce al fatto che le proprietà o gli stati mentali sopravvenienti desumono o ereditano l'efficacia causale dalle loro rispettive basi fisiche di sopravvenienza. Dunque per Kim l'esemplificazione di un certo genere mentale in una data occasione, in virtù del suo realizzatore fisico, eredita e conserva i poteri causali del suo realizzatore mediante la relazione di sopravvenienza.

Di tutt'altro avviso abbiamo visto invece essere Putnam, il quale sposa un'idea di "sopravvenienza globale" che non si riferisce alla correlazione tra stati mentali e stati fisici, bensì all'idea che tutte le nostre capacità e abilità di esseri umani sopravvengono sugli stati dell'universo fisico.

Come è emerso dalla discussione, la visione di Putnam del mentale è in profondo disaccordo con la visione di Kim. Putnam infatti rifiuta l'idea che uno stato mentale sia qualcosa che sta dentro le nostre teste, qualcosa che potrebbe essere spiegabile attraverso la ricerca scientifica. In Putnam è sempre presente la convinzione, sin da quando elaborò la teoria causale del riferimento, che le nostre relazioni con l'ambiente esterno sono fondamentali nel dare un contenuto ai nostri stati mentali, e che la mente non è un oggetto che sta dentro il cervello, come potrebbe lasciar intendere una caratterizzazione del mentale come quella offerta da Kim. Secondo Putnam invece quando si parla di mente si sta parlando di un insieme di capacità e di abilità possedute dagli esseri umani, che sono in costante relazione con l'ambiente nel quale viviamo. Nella nozione di sopravvenienza proposta da Putnam ritroviamo queste sue osservazioni sulla natura del mentale e sul suo ruolo nel caratterizzare la nostra vita.

Da ultimo abbiamo illustrato la visione di White. Questa ci è parsa particolarmente significativa poiché sembra racchiudere e riassumere in sé il discorso qui affrontato. White infatti ci parla di una impossibilità di principio della relazione tra mentale e fisico, e la soluzione del dilemma di come queste due sfere interagiscano non sta, come per esempio sostenevano i riduzionisti più estremi, nel fatto che il mentale è uguale al fisico. La sua soluzione va nella direzione opposta, ovvero nel vedere mentale e fisico completamente

diversi e imparagonabili. Se le cose stanno così, il problema di come possano interagire tra loro neanche si pone.

Parrebbe quindi che una possibile soluzione al problema mente-corpo non sia la sopravvenienza "debole" di Davidson, né la sopravvenienza "forte" di Kim e neanche la sopravvenienza "globale" di Putnam o la non sopravvenienza *tout court*. La soluzione potrebbe proprio essere la *a-sopravvenienza* proposta da White.

La riflessione di White su tale argomento pare fornirci una nuova consapevolezza: il fatto che per tutto questo tempo i filosofi non abbiano trovato un accordo sulla nozione di sopravvenienza e su come intenderla, come abbiamo cercato di mostrare nella ricerca, forse può significare che la questione è davvero irrisolvibile, forse la soluzione da attuare è proprio quella proposta da White, cioè quella di mantenere le due dimensioni separate, e non porsi neanche il problema di come queste interagiscono perché per natura non ha senso. Ciò che White afferma sostanzialmente è che mancano i presupposti per porsi il problema mente-corpo poiché stiamo parlando di due ambiti incommensurabili, ossia di due dimensioni imparagonabili, senza alcuna possibilità di trovare un rapporto tra di esse.

Le differenze nel concepire la nozione di sopravvenienza, oltre a essere dovute alla diversa concezione degli stati mentali e della realtà che hanno gli autori di cui abbiamo discusso, sono dovute anche alla diversa caratterizzazione del concetto di causa. Kim, per esempio, insiste molto su questo punto nella sua critica al monismo anomalo: tale critica infatti, come abbiamo più volte sottolineato, si basa proprio sull'assenza di un ruolo causale delle proprietà mentali. Così come importante è tale concetto nella critica della Hornsby: anche lei pone l'accento sulle criticità della caratterizzazione della causalità in Davidson. Parlando di Putnam abbiamo notato invece che egli sposa una nozione di causa diversa da quella intesa da Kim e Hornsby. Putnam ritiene che il concetto di causazione si trova in rapporto di interdipendenza con quello di *spiegazione*: dunque spiegazioni differenti portano a diverse concezioni della causalità, tutte ugualmente legittime.

Al riguardo, ci pare che la soluzione proposta da White superi anche questo problema della causalità, ovvero quale sia il rapporto causale tra mentale e fisico. Ancora una volta infatti in White non si pone il problema se è il fisico a causare il mentale, o se il mentale conserva una qualche autonomia causale rispetto al fisico, poiché essendo le due dimensioni incommensurabili, il problema della causalità del mentale posto in questi termini non ha motivo di sussistere. Anzi, proprio nel concepire il mentale come dimensione

incommensurabile rispetto alla dimensione fisica, White rende giustizia a tutto ciò che concerne il mentale.

E si noti che non è un mantenere le due dimensioni separate *à la* Descartes: il mentale in White non è una sostanza a sé, la a-sopravvenienza non implica che le proprietà mentali aleggino libere dal fisico e che la mente sia, come voleva Descartes, una sostanza separata dal mondo fisico, priva di dimensione temporale, indefinita, libera e inestesa. White ritiene che vedere mentale e fisico come sostanze separate porterebbe all'idea che il mentale possa aleggiare libero dal fisico (o viceversa), idea che egli non accetta. Come abbiamo sottolineato nel quarto capitolo, il dualismo che scaturirebbe dalla visione di White sembrerebbe più un dualismo delle proprietà, che vede appunto proprietà mentali e proprietà fisiche separate e irriducibili l'una all'altra.

La proposta avanzata da quest'ultimo su come intendere la questione della relazione tra mentale e fisico, lo ribadiamo, ci pare racchiuda in sé il discorso che abbiamo affrontato in questa tesi, ossia ci pare dare una risposta convincente alle difficoltà poste dal problema mente-corpo e dalla nozione di sopravvenienza che abbiamo discusso. La morale dunque che possiamo trarre da decenni di discussioni sulla sopravvenienza, legate a quale sia il modo migliore di intenderla e a che tipo di relazioni essa dia luogo, e a tutte le altre problematiche delle quali si è cercato di dare una illustrazione, ci sembra proprio quella proposta da White: il problema non sussiste poiché non c'è una base comune tra mentale e fisico sulla quale poter argomentare che tipo di relazione esista tra essi, e questo perché le due sfere sono incommensurabili.

BIBLIOGRAFIA

Antonietti, A.
1960 *Il luogo della mente*, Franco Angeli, Milano.

Antony, M. V.
2003 *Davidson's Argument for Monism*, "Synthese", Vol. 135, No. 1, pp. 1-12.

Baghramian, M.
2008 *"From Realism Back to Realism": Putnam's Long Journey*, "Philosophical Topics", Vol. 36, No. 1, pp. 17-35.
2013 *A Life in Philosophy*, in Baghramian 2013, pp. 1-16.

Baghramian, M. (a cura di)
2013 *Reading Putnam*, Routledge, Oxon.

Balog, K.
2012 *Acquaintance and the Mind-Body Problem*, in Gozzano e Hill 2012, pp. 16-42.

Baker, L. R.
1992 *Metaphysics and Mental Causation*, in Heil e Mele 1992, pp. 75-95.

Beakley, B. e Ludlow, P. (a cura di)
2006 *The Philosophy of Mind: Classical Problems/Contemporary Issues*, MIT Press, Cambridge (Mass.).

Benbaji, H.
2010 *Token Monism, Event Dualism and Overdetermination,* "Canadian Journal of Philosophy", Vol. 40, No. 1, pp. 63-82.

Ben-Menahem, Y. (a cura di)
2005 *Hilary Putnam*, Cambridge University Press, Cambridge, New York.

Bickle, J.
2012 *A Brief History of Neuroscience's Actual Influences on Mind-Brain Reductionism*, in Gozzano e Hill 2012, pp. 88-110.

Biletzki, A. (a cura di)
2010 *Hues of Philosophy – Essay in Memory of Ruth Manor*, College Publications, London.

Block, N.
1980a (a cura di) *Readings in Philosophy of Psychology*, Harvard University Press, Cambridge (Mass.).
1980b Introduction: What is Functionalism?, in Block 1980a, pp. 171-184.

Bonomi, A. (a cura di)
2001 *La struttura logica del linguaggio*, Bompiani, Milano.

Brown, R. e Ladyman, J.
2009 *Physicalism, Supervenience and the Fundamental Level*, "The Philosophical Quarterly", Vol. 59, No. 234, pp. 20-38.

Burge, T.
1979 *Individualism and the Mental*, in Paternoster 1999, pp. 91-128.
2007 *Foundations of Mind*, Clarendon Press, Oxford.

Caston, V.
1997 *Epiphenomenalisms, Ancient and Modern*, "The Philosophical Review", Vol. 106, No. 3, pp. 309-363.

Chakraborti, C.
2005 *Mental Properties and Levels of Properties*, "Metaphysica", Vol. 6, No. 2, pp. 7-24.

Chalmers, D.
1996 *The Conscious Mind: in Search of a Fundamental Theory*, Oxford University Press, New York.

Churchland, P.
1989 *A Neurocomputational Perspective. The Nature of Mind and the Structure of Science*, MIT Press, Cambridge (Mass.), trad. it. Il Mulino, Bologna, 1992.
1990 *Could a Machine Think?*, in "Scientific American", vol. 262, pp. 32-37.
1995 *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge (Mass.), trad. it. Il Saggiatore, Milano, 1998.

Code, A.
1991 *Aristotle, Searle, and the Mind-Body Problem*, in Lepore e Van Gulick 1991, pp. 105-113.

Coliva, A.
2007 Lo scetticismo sull'esistenza del mondo esterno, in De Caro e Spinelli 2007, pp. 185-209.

Conant, J. e Żegleń, U. M. (a cura di)
2002 *Hilary Putnam: Pragmatism and Realism*, Routledge, Londra.

Crane, T.
1995 *The Mental Causation Debate,* " Proceedings of the Aristotelian Society Supplementary", Vol.69, pp. 211-253.
2001 *Elements of Mind,* Oxford University Press, Oxford, trad. it. Cortina, Milano, 2003.

Davidson, D.
1980 *Essays on Actions and Events*, Oxford University Press, New York, trad. it. Il Mulino, Bologna, 1992.
1986 *Knowing One's Own Mind*, in Paternoster 1999, pp. 175-198.
1992 *Thinking Causes*, in Heil e Mele 1992, pp. 3-17.
1994 *Self Portrait*, in Guttenplan 1994, pp. 231-236.
1999 *Reply to Jennifer Hornsby*, in Hahn 1999, pp. 636-640.

De Caro, M.
1996a *Epistemologia e interpretazione: l'esternalismo di Donald Davidson*, "Rivista Filosofica", Vol. 87, pp. 315-341.
1996b *Il lungo viaggio di Hilary Putnam*, "Lingua e Stile", Vol. 31, No. 4, pp. 527-545.
1997 *Monismo anomalo ed epifenomenismo*, "Il Cannochiale", Vol. 11, pp. 255-267.
1998 *Dal punto di vista dell'interprete: la filosofia di Donald Davidson*, Carocci, Roma.
1999 *Davidson in Focus*, in De Caro (a cura di) 1999, pp. 1-29.
2003 *Responsabilità e scetticismo*, in Egidi, Dell'Utri e De Caro 2003, pp. 31-40.
2004a *Il libero arbitrio: una introduzione*, Laterza, Bari.
2004b *Is Freedom Really a Mystery*, in De Caro e Macarthur 2004, pp. 183-195.
2007 *How to Deal with the Free Will Issue: the Role of Conceptual Analysis and Empirical Science*, in Marraffa, De Caro e Ferretti 2007, pp. 255-270.
2008 *Azione*, il Mulino, Bologna.
2011 *Autonomia della filosofia e neuroscienze*, "Rivista di filosofia", Vol. 102, pp. 278-283.

De Caro, M. (a cura di)
1999 *Interpretations and Causes. New Perspectives on Donald Davidson's Philosophy*, Kluwer, Dordrech.

De Caro, M; Lavazza, A. e Sartori, A. (a cura di)
2010 *Siamo davvero liberi? Le neuroscienze e il mistero del libero arbitrio*, Codice Edizioni, Torino.

De Caro, M. e Macarthur, D. (a cura di)
2004 *Naturalism in Question*, Harvard University Press, Cambridge (Mass.), trad. it. Fazi, Roma, 2005.
2010 *Naturalism and Normativity*, Columbia University Press, New York.
2012 *Hilary Putnam: Artisanal Polymath of Philosophy*, in Putnam 2012a, pp. 7-52.

De Caro, M. e Maffettone, S.
2000 *Cosa dobbiamo intendere come persona : ragioni del corpo, ragioni della mente*, "Iride", Vol. 13, No. 31, pp. 549 – 561.

De Caro, M. e Spinelli, E. (a cura di)
2007 *Scetticismo: una vicenda filosofica*, Carocci, Roma.

Dell'Utri, M.

1992 *Le vie del realismo*, Franco Angeli, Milano.
2007 *Scetticismo e fallibilismo,* in De Caro e Spinelli 2007, pp. 149-168.
2009 *Conoscenza umana tra verità e scetticismo: nuove idee su un vecchio problema*, http://www.treccani.it/enciclopedia/conoscenza-umana-tra-verita-e-scetticismo_%28XXI-Secolo%29.

Dennett, D. C.
1997 *Kinds of Minds*, Weidenfeld & Nicolson, London, trad. it. Sansoni, Milano, 1997.

De Palma, A. e Pareti, G. (a cura di)
2004 *Mente e corpo: dai dilemmi della filosofia alle ipotesi della neuroscienza*, Boringhieri, Torino.

Descartes, R.
1641 *Meditationes de Prima Philosophia,* trad. it. Laterza, Bari, 1954.

Devitt, M.
1997 *Realism and Truth*, Princenton University Press, New Jersey.

Di Francesco, M.
1996 *Introduzione alla filosofia della mente*, Nuova Italia Scientifica, Roma.
1998 *La mente e le sue rivoluzioni*, www.swif.uniba.it.
1998 *L'io e i suoi sé: identità personale e scienza della mente*, Cortina, Milano.
2004 *Filosofia della mente*, V. 1.0, in L. Floridi (a cura di), *Linee di ricerca*, SWIF, pp. 472-522.

Di Lorenzo Ajello, F.
1998 *Mente, azione e linguaggio nel pensiero di John R.. Searle*, Franco Angeli, Milano.

Dreyfus, H.
1972 *What Computers Can't Do: The Limits of Artificial Intelligence*, Harper and Row, New York, trad. it. Armando, Roma, 1988.

Eccles, J. e Robinson, N.
1984 *The Wonder of Being Human*, Free Press, New York, trad. it. Armando, Roma, 1985.

Egidi, R.; Dell'Utri, M. e De Caro, M.
2003 *Normatività fatti valori*, Quodlibet, Macerata.

Ehring, D.
1996 *Mental Causation, Determinables and Property Instances*, "Noûs", Vol. 30, No. 4, pp. 461-480.

Evnine, S.
1991 *The Philosophy of Donald Davidson*, Stanford University Press, Stanford (California).

Feigl, H.
1967 *The "Mental" and the "Physical": The Essay and a Postscript*, University of Minnesota Press, Minneapolis.

Ferraris, M.
2006 *Un "cocktail" di Searle*, http://www.rescogitans.it.

Ferretti, F. (a cura di)
2001 *Jerry A. Fodor, mente e linguaggio*, Laterza, Roma-Bari.

Feyerabend, P.
1975 *Against Method. Outline of an Anarchistic Theory of Knowledge*, Verso, London, trad. it. Feltrinelli, Milano, 1979.

Floyd, J.
2009 *Recent Themes in the History of Early Analytic Philosophy*, "Journal of the History of Philosophy", vol. 47, No. 2, pp. 157–200.

Fodor, J.
1985 *Fodor's Guide to Mental Representation: The Intelligent Auntie's Vademecum*, in Paternoster 1999, pp. 76-100.
2000 *The Mind Doesn't Work That Way*, MIT Press, Cambridge (Mass.), trad. it. Laterza, Roma-Bari, 2001.

Fornero, G. e Tassinari, S.
2002 *Le filosofie del novecento, vol. 2*, Bruno Mondadori, Milano.

Freeman, C., e Skarda, A.
1991 *Mind/Brain Science: Neuroscience on Philosophy of Mind*, in Lepore e Van Gulick 1991, pp. 115-127.

Frege, G.
1892 *Über Sinn und Bedeutung*, in Bonomi 2001, pp. 9-32.

Friggieri, J.
1981 *Linguaggio e azione. Saggio su J.L. Austin*, Vita e Pensiero, Milano.
1991 *Actions and Speech Actions in the Philosophy of J.L Austin*, Mireva Publications, Malta.

Garfield, J. L.
1988 *Belief in Psychology*, MIT Press, Cambridge, (Mass.), trad. it., il Mulino, Bologna, 1994.

Giorello, S. e Strata P.
1991 *L'automa spirituale: menti, cervelli e computer*, Laterza, Roma-Bari.

Gozzano, S.
1999 *Davidson on Rationality and Irrationality*, in De Caro 1999, pp. 137-149.
2003 *Second Order Properties: Why Kim's Reduction Does Not Work*, "Logic and Philosophy of Science", Vol. 1, No. 1, pp. 1-15.
2007 *Pensieri materiali. Corpo, mente e causalità*, De Agostini Scuola, Novara.
2008 *Tropes' Simplicity and Mental Causation*, in Gozzano e Orilia 2008, pp. 133-154.
2012 *Type-Identity Conditions for Phenomenal Properties*, in Gozzano e Hill 2012, pp. 111-126.

Gozzano, S. e Hill, C. S. (a cura di)
2012 *New Perspectives on Type Identity: the Mental and the Physical*, Cambridge University Press, Cambridge (Mass.).

Gozzano, S. e Orilia, F. (a cura di)

2008 *Tropes, Universals and the Philosophy of Mind*, Ontos, Francoforte.

Guttenplan, S. (a cura di)
1994 *A Companion to the Philosophy of Mind*, Blackwell, Oxford.

Hahn, L. E. (a cura di)
1999 *The Philosophy of Donald Davidson*, Open Court, Chicago.

Hardcastle, V. G.
1998 *On the Matter of Minds and Mental Causation*, "Philosophy and Phenomenological Research", Vol. 58, No. 1, pp. 1-25.

Heckmann, H. D.
1992 *Mental Events Again: or What Is Wrong with Anomalous Monism?,* "Erkenntnis", Vol. 36, No. 3, pp. 345-373.

Heil, J.
1999 *Multiple Realizability*, "American Philosophical Quarterly", Vol. 36, No. 3, pp. 189-208.

Heil, J. e Mele, A.
1992 *Mental Causation*, Clarendon Press, Oxford.

Heil, J. e Robb, D.
2003 *Mental Properties*, "American Philosophical Quarterly", Vol. 40, No. 3, pp. 175-196.

Herstein, G. L.
2005 *Davidson on the Impossibility of Psychophysical Laws*, "Synthese", Vol. 145, No.1, pp. 45–63.

Hofstadter D. R. e Dennett D. C.
1981 *The Mind's I*, Basic Book, New York, trad. it. Adelphi, Milano, 1993.

Honderich, T.
1982 *The Argument for Anomalous Monism*, "Analysis", Vol. 42, No. 1, pp. 59-64.

Horgan, T.
1993 *From Supervenience to Superdupervenience: Meeting the Demands of a Material World*, "Mind", Vol. 102, No. 408, pp. 555-586.
1996 *Kim on the Mind-Body Problem*, "The British Journal of Philosophy of Science", Vol. 47, No. 4, pp. 579-607.
1997 *Kim on Mental Causation and Causal Exclusion*, "Noûs, Vol. 31, "Philosophical Perspectives", 11, pp. 165-184.

Hornsby J.
1980-1 *Which Physical Events Are Mental Events?,* "Proceedings of the Aristotelian Society", New Series, Vol. 81, pp. 73-92.
1985 *Physicalism, Events, and Part-Whole Relations*, in Vermazen e Hintikka 1985, pp. 444-458.
1999 *Anomalousness in Action*, in Hahn (Ed) 1999, pp. 623-635.
2004 *Agency and Alienation*, in De Caro e Macarthur 2004, pp. 167-182.

Horowitz, A.

1999 *Is There a Problem in Physicalist Epiphenomenalism?*, "Philosophy and Phenomenological Research", Vol. 59, No. 2, pp. 421-434.

Jackson, F. e Pettit, P.
1900 *Causation in the Philosophy of Mind*, "Philosophy and Phenomenological Research", Vol. 50, pp. 195-214.

Jaworski, W.
2006 *Mental Causation from the Top-Down*, "Erkenntnis", Vol. 65, No. 2, pp. 277-299.

Johnson D. M. e Erneling C. E. (a cura di)
1997 *The Future of the Cognitive Science*, Oxford University Press, New York.

Johnson-Laird, P.
1993 *The Computer and the Mind. An Introduction to Cognitive Science*, W. Collins Sons & Co., London, trad. it. il Mulino, Bologna, 1997.

Keaton, D.
2010 *Two Kinds of Role Property*, "Philosophia", Vol. 38, pp. 773-788.

Kim, J.
1978 *Supervenience and Nomological Incommensurables*, "American Philosophical Quartely", Vol. 15, No. 2, pp. 149-156.
1985 *Psychophysical Laws*, in Vermazen e Hintikka 1985, pp. 369-386.
1989 *The Myth of Nonreductive Materialism*, "Proceedings and Addresses of the American Philosophical Association", Vol. 63, No. 3, pp. 31-47.
1992 *Can Supervenience and "Non-Strict Laws" Save Anomalous Monism?* in Heil e Mele 1992, pp. 19-26.
1993a *Mental Causation in a Physical World*, "Philosophical Issues", Vol.3, pp. 157-176.
1993b *Supervenience and Mind*, Cambridge University Press, Cambridge.
1995 *Mental Causation: What? Me Worry?*, "Philosophical Issues", Vol. 6, Content, pp. 123-151.
1998 *Mind in a Physical World*, MIT Press, Cambridge (Mass.).
2003 *Philosophy of Mind and Psychology*, in Ludwig 2003, pp. 113-136.
2005 *Physicalism, or something near enough*, Princeton University Press, Princeton.
2006 *Multiple Realization and the Metaphysics of Reduction*, in Beakley e Ludlow 2006, pp. 177-199.
2012 *The Very Idea of Token Physicalism*, in Gozzano e Hill 2012, pp. 167-185.

Kistler, M.
1999 *Multiple Realization, Reduction and Mental Properties*, "International Studies in the Philosophy of Science", Vol. 13, No. 2, pp. 135-149.

Klagge, J. C.
1990 *Davidson's Troubles with Supervenience*, "Synthese", Vol. 85, No. 2, pp. 339-352.

Kripke, S.
1972 *Name and Necessity*, Blackwell, Oxford, trad. it. Borighieri, Torino, 1999.

Kuhn, T. S.

1982 *Commensurability, Comparability, Communicability*, PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 2: Symposia and Invited Papers, pp. 669-688.

Latham, N.
1999 *Davidson and Kim on Psychophysical Laws*, "Synthese", Vol. 118, No. 2, pp. 121–143.

Laudisa, F.
2000 *Causalità*, Carocci, Roma.

Laurier, D.
2005 *Mind, Davidson and Reality*, "Principia", Vol. 9, No. 1-2, pp. 125-157.

Leonardi, P.
1999 *Anomalous Monism*, in De Caro 1999, pp. 117-124.

Lepore, E. e Loewer, B.
1987 *Mind Matters*, "The Journal of Philosophy", Vol. 84, No. 11, pp. 630-642.

Lepore, E. e McLaughlin (a cura di)
1985 *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, New York.

Lepore, E. e Van Gulick R. (a cura di)
1991 *John Searle and His Critics*, Blackwell, Cambridge.

Lewis, D.
1986 *On the Plurality of Worlds*, Blackwell, Oxford.

Ludwig, K. (a cura di)
2003 *Donald Davidson: Contemporary Philosophy in Focus*, Cambridge University Press, Cambridge (Mass.).

Lyons, W. (a cura di)
1995 *Modern Philosophy of Mind*, Everyman, Guernsey.

Macarthur, D.
2004 *Putnam's Natural Realism and the Question of a Perceptual Interface*, "Philosophical Explorations", Vol. 7, No. 2, pp. 159-173.

Marconi, D.
2001 *Filosofia e scienza cognitiva*, Laterza, Roma-Bari.

Marcus, E.
2005 *Mental Causation in a Physical World*, "Philosophical Studies", Vol. 122, No. 1, pp. 27-50.

Marraffa, M.
2007 *Setting the Stage: Persons, Minds and Brains*, in Marraffa, De Caro e Ferretti 2007, pp. 3-22.

Marraffa, M; De Caro, M. e Ferretti.F. (a cura di)
2007 *Cartographies of the Mind. Philosophy and Psychology in Intersection*, Springer,

Dordrecht.

Marras, A.
1994 *Nonreductive Materialism and Mental Causation*, "Canadian Journal of Philosophy", Vol. 24, No. 3, pp. 465-493.
2001 *On Putnam's Critique of Metaphysical Realism: Mind-Body Identity and Supervenience*, "Synthese", Vol. 126, No. 3, pp. 407-426.
2002 *Kim on Reduction*, "Erkenntnis", Vol. 57, No. 2, pp.231-257.
2007 *Kim's Supervenience Argument and Nonreductive Physicalism*, "Erkenntnis", Vol. 66, No. 3, pp. 305-327.

Marras , A. e Yli-Vakkuri, J.
2008 *The "Supervenience Argument": Kim's Challenge to Nonreductive Physicalism*, in Gozzano e Orilia 2008, pp. 101-132.

Marzano, M. M.
2003 *The Ontology of Value*, in Egidi, Dell'Utri e De Caro 2003, pp. 61-70.

McDowell, J.
1994 *Mind and World*, Harvard University Press, Massachusetts.
1998 *Meaning, Knowledge and Reality*, Harvard University Press, Cambridge (Mass.).

McIntyre, L.
1998 *Davidson and Social Scientific Laws*, "Synthese", Vol. 120, No. 3, pp. 375-394
2002 *Supervenience and Explanatory Exclusion*, "Crítica", Vol. 34, No. 100, pp. 87-101.

McLaughlin, B.
1992 *On Davidson's Response to the Charge of Epiphenomenalism*, in Heil e Mele 1992, pp. 27-40.

Melchert, N.
1986 *What's Wrong with Anomalous Monism?*, "The Journal of Philosophy", Vol. 83, No. 5, pp. 265-274.

Menzies, P. e Price, H.
1993 *Causation as a Secondary Quality*, "The British Journal of the Philosophy Science", Vol. 44, No.2, pp. 187-203.

Moore, G.E.
1912 *Ethics*, Cambridge University Press, Cambridge.

Moravia, S.
1987 *L'enigma della mente. Il "Mind-Body" problem nel pensiero contemporaneo*, Laterza, Roma.

Morowitz, H. J.
1980 *Rediscovering the Mind*, in Hofstadter e Dennett 1981, pp. 44-52.

Moser, P. K. e Trout, J. D. (a cura di)
1995 *Contemporary Materialism: a Reader*, Routledge, London.

Nagel, T.

1974 *What Is It Like to Be a Bat?*, in Hofstadter e Dennett 1981, pp. 379-392.
1991 *Coscienza e realtà oggettiva*, in Giorello e Strata 1991, pp. 31-37.

Nannini, S.
1999 *Physicalism and the Anomalism of the Mental*, in De Caro 1999, pp. 101-116.
2002 *L'anima e il corpo. Un'introduzione storica alla filosofia della mente*, Laterza, Roma-Bari.
2005 *Mente e corpo nella filosofia analitica contemporanea*, in «Laboratorio dell'ISPF», www.ispf.cnr.it/ispf-lab.

Needham, P.
2009 *Reduction and Emergence: A Critique of Kim*, "Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition", Vol. 146, No. 1, pp. 93-116.

Ney, A.
2012 *The Causal Contribution of Mental Events*, in Gozzano e Hill 2012, pp. 230-250.

Orilia, F.
2008 *Basic Ontology, Multiple Realizability and Mental Causation*, in Gozzano e Orilia 2008, pp. 57-100.

Owens, J.
1993 *Content, Causation, and Psychophysical Supervenience*, "Philosophy of Science", Vol. 60, No. 2, pp. 242-261.

Paternoster, A.
1999 (a cura di) *Mente e linguaggio*, Guerini, Milano.
2002 *Introduzione alla filosofia della mente*, Laterza, Roma-Bari.
2007 *Vision Science and the Problem of Perception*, in Marraffa, De Caro e Ferretti 2007, pp. 53-64.

Perissinotto, L.
2002 *Le vie dell'interpretazione nella filosofia contemporanea*, Laterza, Bari.

Perry, J.
2012 *Return of the Zombies?*, in Gozzano e Hill 2012, pp. 251-263.

Pessa, E. e Penna, M. P.
1997 *Manuale di scienza cognitiva: intelligenza artificiale classica e psicologia cognitiva*, Laterza, Roma-Bari.

Pessin, A. e Goldberg S.
1996 *The Twin Heart Chronicles: Twenty Years of Reflections on Hilary Putnam's "The Meaning of Meaning"*, M. E. Sharpe, New York.

Piccinini, G.
2007 *Computational Explanation and Mechanistic Explanation of Mind*, in Marraffa, De Caro e Ferretti 2007, pp. 23-36.

Place, U.T.
1956 *Is consciousness a brain process?* , British Journal of Psychology, Vol. 47, No.1, pp.44-50.

Popper, K. e Eccles, J.
1977 *The Self and Its Brain: An Argument for Interactionism*, Springer, Berlin, trad. it. Armando, Roma.

Pratt, V.
1987 *Thinking Machines. The Evolution of Artificial Intelligence*, Basil Blackwell, Oxford, trad. it., Il Mulino, Bologna, 1990.

Putnam, H.
1975a *Mind, Language and Reality*, *Philosophical Paper*, Vol. 2, Cambridge University Press, Cambridge (Mass.), trad. it., Adelphi, Milano, 1987.
1975b *Introduction*, in Putnam 1975, trad. it., pp. 11-21.
1975c *The Meaning of 'Meaning'*, cap. 12, in Putnam 1975a, pp. 239-297.
1975d *Philosophy and our Mental Life*, cap. 14, in Putnam 1975a, trad. it. pp. 318-330.
1975e *Brains and Behavior*, cap. 16, in Putnam 197a5, trad. it., pp. 354-371.
1975f *Minds and Machines*, cap. 18, in Putnam 1975a, trad. it., pp. 392-415.
1975g *The Mental Life of Some Machines*, cap. 20, in Putnam 1975a, trad. it., pp. 439-460.
1975h *The Nature of Mental States*, cap. 21, in Putnam 1975a, trad. it., pp. 461-472.
1982 *Why There Isn't a Ready-Made World*, "Synthese", Vol. 51, No. 2, pp. 205-228, ristampato in Putnam 1983, pp. 205-228.
1983 *Realism and Reason, Philosophical Papers*, Vol. 3, Cambridge University Press, Cambridge (Mass.).
1987 *The Many Faces of Realism*, Open Court Publishing Company, La Salle, Illinois, trad. it., Garzanti, Milano, 1991.
1988 *Representation and Reality*, MIT Press, Cambridge (Mass.), trad. it. Garzanti, Milano,1993.
1990 *Realism with a Human Face*, Harvard University Press, Cambridge (Mass.), trad. it. Il Mulino, Bologna, 1995.
1992 *Renewing Philosophy*, Harvard University Press, Cambridge (Mass.).
1994 *Words and Life*, Harvard University Press, Cambridge.
1996 *Introduction*, in Pessin e Goldberg 1996, pp. xv-xii.
1997a *Functionalism: Cognitive Science or Science Fiction?*, in Johnson e Erneling 1997, pp. 32-44, ristampato in Putnam 2012a, pp. 608-623.
1997b *A Half Century of Philosophy, Viewed from Within*, "Daedalus", Vol. 126, No. 1, pp. 175-208.
1999 *The Threefold Cord: Mind, Body and World*, Columbia University Press, New York, trad. it. Il Mulino, Bologna, 2003.
2002 *Comment on Tadeusz Szubka's paper*, in Conant e Żegleń 2002, pp. 125-127.
2004 *The Content and Appeal of "Naturalism"*, in De Caro e Macarthur 2004, pp. 45-57.
2008 *Reply to Stephen White*, "European Journal of Analytic Philosophy", Vol. 4, No. 2, pp. 29-32.
2010 *Science and Philosophy*, in De Caro e Macarthur 2010, pp. 89-99.
2012a *Philosophy in an Age of Science: Physics, Mathematics, and Skepticism*, Harvard University Press, Cambridge (Mass.), trad. it., Il Mulino, Bologna, 2012.
2012b *Corresponding with Reality*, cap. 3, in Putnam 2012a, pp. 97-120.
2012c *On Not Writing and Appeal of "Naturalism"*, cap. 4, in Putnam 2012a, pp. 121-142.
2012d *The Content and Appeal of "Naturalism"*, cap. 5, in Putnam 2012a, pp. 109-125.
2012e *Wittgenstein: A Reappraisal*, cap. 28, in Putnam 2012a, pp. 223-236.
2012f *Aristotle's Mind and the Contemporary Mind*, cap.34, in Putnam 2012a, pp. 275-302.
2012g *How to Be a Sophisticated "Naïve Realist"*, cap. 36, in Putnam 2012a, pp. 303-321.

Rainone, A.

1999 *Thirty-Five Years after "Action, Reasons, and Causes": What has become of Davidson's Causal Theory of Action?,* in De Caro 1999, pp. 125-136.

Rapaport, W. J.
1995 *Understanding Understanding: Syntatic Semantics and Computational Cognition*, in Tomberlin, AI, Connectionism, and Philosophical Psychology, *Philosophical Perspectives*, vol. 9, http://www.cse.buffalo.edu.

Rieber, S.D.
2003 *Causation as Property Acquisition*, "Philosophical Studies", Vol. 109, No. 1, pp. 53-74.

Robb, D.
1. *The Properties of Mental Causation*, "The Philosophical Quarterly", Vol. 47, No.187, pp. 178-194.

Rorty, R.
1979 *Philosophy and the Mirror of Nature*, Princeton University Press, Princeton, trad. it. Bompiani, Milano, 1998.

Rosenthal, D.
2010 *How to Think About Mental Qualities*, "Philosophical Issues", Vol. 20, pp. 368–393.

Ryle, G.
1949 *The Concept of Mind*, Barnes & Noble, New York, trad. it. Laterza, Roma-Bari, 2007.

Shagrir, O.
2005 *The Rise and Fall of Computational Functionalism*, in Ben-Menahem 2005, pp. 220-250.
2009 *Anomalism and Supervenience: a Critical Survey*, "Canadian Journal of Philosophy", Vol. 39, No. 2, pp. 237-272.
2010 *Davidson's Notion of Supervenience*, in Biletzki, pp. 43-58.
2011 *Supervenience and Anomalism are Compatible*, "Dialectica", Vol. 65, No. 2, pp. 241-266.

Shapiro, L. A.
2010 *Lessons from Causal Exclusion*, "Philosophy and Phenomenological Research", Vol. 81, No. 3, pp. 594-604.

Searle, J. R.
1980 *Minds, Brains, and Programs*, in Hofstadter e Dennett 1981, pp. 341-360.
1983 *Intentionality, An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge, trad. it. Bompiani, Milano, 1985.
1984 *Minds, Brains and Science*, Harvard University Press, Cambridge (Mass.), trad. it. Bompiani, Milano, 1987.
1991 *Response*, in Lepore e Van Gulick 1991.
1992 *The Rediscovery of the Mind*, MIT Press, Cambridge (Mass.), trad. it. Boringhieri, Torino, 1994.

Smart, J.J.C.
1959 *Sensations and Brain Processes*, "Philosophical Review", Vol. 68, No. 2. pp. 141-156.

Sober, E.
1999 *Physicalism from a Probabilistic Point of View*, "Philosophical Studies", Vol. 95, No. 1-

2, pp. 135-174.

Sosa, E.
1993 *Davidson's Thinking Causes*, in Heil. e Mele 1992, pp. 41-50.

Stalnaker, R.
1996 *Varieties of Supervenience*, Noûs, Vol. 30, "Philosophical Perspectives", 10, Metaphysics, pp. 221-241.

Szubka, T.
2002 *The Causal Theory of Perception and Direct Realism*, in Conant e Żegleń 2002, pp. 109-124.

Swoyer, C. e Orilia, F.
2011 *Properties*, The Stanford Encyclopedia of Philosophy (Winter 2011 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/win2011/entries/properties/.

Tomberlin, J. E.
1995 *Mental Causation: a Query for Kim*, "Philosophical Issue", Vol. 6, Content, pp. 152-159.

Turing, A. M.
1950 *Computing Machinery and Intelligence*, in Hofstadter e Dennett, 1981, pp. 61-75.

Van Brakel, J.
1999 *Supervenience and Anomalous Monism*, "Dialectica", Vol. 53, No. 1, pp. 3-25.

Varzi, A.
2011 *Mereology*, The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2011/entries/mereology/.

Vermazen B. e Hintikka M. B.
1985 *Essay on Davidson: Actions and Events*, Clarendon Press, Oxford.

Vicente, A.
2001 *El Principio del Cierre Causal del Mundo Físico*, "Crítica", Vol. 33, No. 99, pp.3-17.

White, S.
2004a *Skepticism, Deflation, and the Rediscovery of the Self*, The Monist, 87, pp. 275-298.
2004b *Subjetctivity and the Agential Perspective*, in De Caro e Macarthur 2004, pp. 196-223.
2007a *Empirical Psychology, Transcendental Phenomenology, and the Self*, in Marraffa, De Caro e Ferretti 2007, pp. 243-254.
2007b *The Transcendental Significance of Phenomenology*, "Psyche", Vol. 13, pp. 1-31.
2008 *On the Absence of an Interface: Putnam, Direct Perception, and Frege's Constraint*, "European Journal of Analytic Philosophy", Vol. 4, No. 2, pp. 11-28.
2010a *Risposta di Stephen White a Hilary Putnam*, "Rivista di estetica", Vol. 2, No. 44, pp. 193-203.
2010b *Phenomenology and the Normativity of Practical Reason*, in De Caro e Macarthur 2010, pp. 205-225.

Wiezenbaum, J.
1976 *Computer Power and Human Reason: From Judgment to Calculation*, Freeman and Co., San Francisco, trad. it. Gruppo Abele, Torino, 1987.

Yablo, S.
1992 *Mental Causation*, Philosophical Review, Vol. 101, No. 2, pp. 245-280.
1997 *Wide Causation*, "Noûs, Supplement: Philosophical Perspectives", Vol. 11, pp. 251-281.

Yalowitz, S.
2008 *Anomalous Monism*, The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2008/entries/anomalous-monism.

Yates, D.
2009 *Emergence, Downwards Causation and the Completeness of Physics*, "The Philosophical Quarterly", Vol. 59, No. 234, pp. 110-131.

Ziff, P.
1966 *Philosophic Turnings. Essay in Conceptual Appreciation*, Cornell University Press, Ithaca, trad. it. Laterza, Bari, 1969.

Ringraziamenti

Il primo ringraziamento va al Prof. Massimo Dell'Utri per la fiducia, il supporto, la gentilezza e la costante disponibilità nell'avermi aiutata a realizzare questa tesi.

Un altro doveroso ringraziamento va al Prof. Joe Friggieri per avermi messo a completa disposizione il suo tempo, le sue conoscenze e le risorse sia sue che dell'Università di Malta.

Un grazie anche a mia madre, per il sostegno in questi tre anni e per la serenità che, sono sicura con tanta fatica, mi ha donato in quest'ultimo anno. Un grazie a mio padre, perché il suo orgoglioso e la sua fierezza nei miei confronti sono stati e saranno sempre per me motivo di conforto e incoraggiamento. Un grazie anche a mio fratello, che per sua fortuna ha subito meno degli altri le fasi isteriche del lavoro, ma non ha mai fatto mancare il suo appoggio.

Un grazie speciale a Thomas, per aver sopportato con grande tenacia e tanto affetto questi tre anni fatti delle suddette fasi isteriche, fatti anche di periodi più o meno lunghi di lontananza, e dei classici alti e bassi che accompagnano queste esperienze. È stato la valvola di sfogo e il rifugio dove ritrovare la carica.

Un ringraziamento va anche alle cugine, Anna Laura e Daniela, sempre interessate all'andamento del mio percorso e sempre presenti per qualsiasi esigenza. Un grazie anche a tutti gli amici e i parenti che mi sono stati vicini in questo periodo.