
FANTASCIENZA E FILOSOFIA DELLA MENTE*

MARCO SALUCCI

Facoltà di Psicologia, Università di Firenze
Liceo Scientifico 'P. Gobetti', Bagno a Ripoli, Firenze

1. Esperimenti mentali

Se vi dicessi che sto per raccontarvi una storia in cui compaiono robot, computer, pipistrelli, scienziati del cervello, zombi ed extraterrestri pensereste che stia per raccontarvi una storia di fantascienza: quelli appena citati sono infatti i personaggi tipici di molte storie di fantascienza. Ma così come può accadere di sbagliarsi nel credere che Chinatown si trovi in Cina perché si incontrano solo cinesi, così vi dico che vi state sbagliando e che non sto per raccontarvi una storia di fantascienza ma di filosofia: pipistrelli, robot, zombi ecc. popolano infatti parte del dibattito filosofico contemporaneo, più esattamente quella parte che va sotto il nome di filosofia della mente.

Nel tentativo di fornire una spiegazione naturale dei fenomeni mentali ogni epoca ha utilizzato le sue metafore. A seconda dello stato delle conoscenze scientifiche e delle disponibilità tecniche l'attività mentale è stata paragonata volta a volta a una secrezione del cervello, al vapore di una locomotiva, a una centrale telegrafica o telefonica, a un computer. Da una certa epoca in poi, tuttavia, il paragone mente/computer ha cominciato a divenire qualcosa di più di una semplice metafora. Dalla metà del XX secolo, quando ancora l'uomo comune chiamava, in modo significativo, «cervelli elettronici» i computer e gli addetti ai lavori «cibernetica» la scienza dell'intelligenza artificiale, tanto il mondo della fantascienza, da sempre dominato dai marziani, quanto il dibattito filosofico sulla mente, da sempre dominato dai concetti di anima e di spirito, hanno cominciato a popolarsi di robot e di computer. Una teoria molto diffusa – il funzionalismo – ha anzi esplicitamente sostenuto che la mente è letteralmente un programma per computer, rendendo così l'ultima metafora un serio programma di ricerca laddove, che io sappia, non è accaduto nulla di simile per la metafora del fumo della locomotiva.

Hilary Putnam, uno dei padri fondatori della contemporanea filosofia della mente (ambito al quale limiterò il più possibile l'esposizione che segue) è colui che più di ogni altro ha immaginato situazioni fantascientifiche per discutere antiche e serissime questioni filosofiche: non si può sperare niente di meglio delle sue parole per entrare nel nostro tema.

Chiunque abbia passato molte ore (ben spese o mal spese?) della propria

fanciullezza a leggere storie di missili e di robot, di androidi e di telepati, di civiltà galattiche e di macchine del tempo, sa benissimo che i robot (macchine ipotetiche che simulano il comportamento umano, spesso con sembianze almeno approssimativamente umane) possono essere buoni o tremendi, amici fedeli dell'uomo o suoi acerrimi nemici. I robot possono essere esemplari o pateticamente buffi: incuterci sbigottita ammirazione coi loro poteri sovrumani [...] oppure divertirci con il loro comportamento ingenuo e sempliciotto. Ci sono stati robot innamorati, robot impazziti [...] robot zelanti fino all'irritazione. Almeno nella letteratura fantascientifica, dunque, un robot può essere 'cosciente', il che significa [...] avere sensazioni, pensieri, atteggiamenti, tratti caratteriali. Ma è veramente possibile? E se lo è, quali sono le condizioni necessarie e sufficienti? E perché noi filosofi dovremmo comunque occuparcene? [...] La mia speranza [è] che il problema della mente delle macchine si dimostrerà in grado di offrire, almeno per qualche tempo, un approccio nuovo e stimolante a questioni decisamente tradizionali di filosofia della mente. [...] Sia Oscar uno di questi robot, e immaginiamo che esso abbia la 'sensazione' del rosso. Oscar ha la sensazione del rosso? [...] Oscar vede qualcosa? Pensa, sente qualcosa? È vivo? È cosciente? [13, pp. 416 e 425]

La fantascienza è il genere cinematografico che si presta maggiormente all'incontro fra cinema e filosofia, il che, ovviamente, vale anche per la fantascienza come genere letterario. Il luogo in cui fantascienza e filosofia si incontrano è quello degli esperimenti di pensiero o esperimenti mentali: essi consistono nel provare a immaginare che cosa accadrebbe se si verificassero certe condizioni così e così. I racconti di fantascienza fanno uso in modo caratteristico di condizioni controfattuali, cioè di condizioni immaginarie in cui qualche aspetto della realtà ordinaria viene significativamente mutato. Sovente la situazione immaginata ha lo scopo di mettere alla prova una teoria o un'ipotesi ed assume così il valore di un esperimento, da cui, appunto, la locuzione «esperimento mentale». In filosofia si fa spesso ricorso a esperimenti mentali poiché la filosofia spesso – a torto o a ragione – prescinde o oltrepassa i confini di ciò che – contingentemente o in linea di principio – è empiricamente sperimentabile. Tuttavia anche la scienza, e più spesso di quanto si pensi, usa la tecnica degli esperimenti mentali. Citerò solo uno dei casi più celebri, ma tanto più significativo in quanto sta esattamente alle origini della scienza moderna ed è stato elaborato precisamente da uno dei suoi padri fondatori: Galileo Galilei. Nel cercare prove per sostenere che un corpo in movimento si muove, in assenza di attrito, all'infinito – e con ciò per distruggere uno dei fondamenti della millenaria fisica aristotelica – Galilei non poteva ricorrere a un esperimento reale, l'unico possibile sarebbe stato un viaggio nello spazio. Egli aveva osservato però che il movimento di una palla su una superficie, a parità di altre condizioni, dura tanto più a lungo tanto più la superficie su cui si muove è levigata. Così si spinse a immaginare che se la superficie fosse stata perfettamente levigata il movimento della palla sarebbe durato all'infinito¹.

2. Da *Star Trek* a Descartes

Da quanto si è detto è facile capire come sia impossibile citare i film di fantascienza che si fondano su esperimenti mentali dal momento che tutti, per definizione, lo fanno. Ma qualcuno andrà comunque citato, almeno allo scopo di documentare quanto vado dicendo.²

Immaginiamo che esista un macchinario in grado di analizzare il mio corpo atomo per atomo e quindi di compilare un manuale di istruzioni e una lista dei materiali necessari per costruire il mio corpo (un manuale e una lista del genere già esistono in natura e sono contenuti nel DNA). Immaginiamo poi che manuale e lista vengano trasmessi via radio all'altro capo del mondo e che là il mio corpo venga ricostruito esattamente come il mio atomo per atomo. La domanda è: il mio «gemello» agli antipodi sono ancora «io»? Voglio dire: io sono «lui»? Drammatizziamo ulteriormente: dopo la «costruzione» del mio gemello antipodiano io vengo distrutto. La domanda è: io sono morto o continuo a vivere nel mio gemello? Risponderete: il gemello è un altro, poiché l'io originale è stato distrutto. Immaginiamo allora che invece di essere distrutto senza troppi scrupoli venga smontato atomo per atomo, imballato in apposite scatole (la scatola degli atomi di carbonio, quella degli atomi di azoto ecc.) e spedito agli antipodi insieme al manuale di istruzioni. Là vengo «rimontato» utilizzando il materiale originario. Sono ancora io? Penso che dovrete rispondere di sì e che sarebbe come risvegliarsi dopo un profondo sonno, un po' come risvegliarsi dopo un'ibernazione. Ma allora vi chiedo: dov'è la differenza fra questa situazione e quella in cui avevo un gemello antipodiano? Solo nel materiale con cui sono stato ricostruito, che, in un caso, è quello originario? Certamente no: non ci può essere nessuna differenza fra due atomi di carbonio, uno trovato qui e l'altro agli antipodi. Quindi se ricostruire me agli antipodi è ricostruire *me*, allora anche ricostruire il mio gemello deve essere ricostruire *me* (oppure ciò che fa la differenza fra le due situazioni è la mera distinzione numerica?)

Il problema filosofico con cui abbiamo avuto appena a che fare è quello dell'identità personale, già discusso da Locke e da Hume nel XVII sec., che nasce dalla grande difficoltà di reperire criteri che permettano di definire l'identità di una persona nel tempo e nello spazio. I film in cui è presente una questione simile sono – l'avrete già indovinato – quelli della serie di *Star Trek* le astronavi della quale sono equipaggiate con una macchina per il teletrasporto³. La macchina per il teletrasporto, immaginata per soddisfare le necessità di spostarsi nello spazio potrebbe però anche permetterci di viaggiare nel tempo: se la macchina fosse conservata per mille anni e ricostruisse un altro me stesso fra mille anni. Ma, un momento ... avremmo allora con ciò risolto anche il problema dell'immortalità? Bhè, sì, fintanto che sarà conservato il manuale d'istruzioni per ricostruire me stesso potrà essere ricostruito dopo ogni mia morte. Mi rimproverate di correre troppo con la fantasia? Accetto tale rimprovero, ma non da tutti: vorrei ricordare, senza nessun intento irriverente, che la lettera originaria del cristianesimo, più che l'immortalità dell'anima, promette la resurrezione dei corpi (e,

mediante questa, la resurrezione della persona e non di un mero zombi). Forse il Dio del cristianesimo provvederà alla resurrezione dei corpi consultando il manuale di istruzioni per la «costruzione» di ciascun essere umano? Vi ricordo che ancora nel XVIII secolo Dio veniva chiamato con appellativi quali «il grande architetto», «il grande ingegnere», «il grande matematico», «il grande orologiaio». Con queste ultime questioni abbiamo menzionato un problema, sì, religioso, ma che è stato a lungo un problema filosofico: per tutti i mille anni in cui filosofia e teologia andavano di pari passo⁴.

Facciamo adesso però un passo verso il tema che ho scelto di trattare più in dettaglio. Racconta un saggio cinese: «Ieri notte Chuang Tzu ha sognato di essere una farfalla svolazzante e ignara di essere Chuang Tzu. Al risveglio si è accorto con stupore di essere Chuang Tzu. Adesso non sa più se è Tzu che ha sognato di essere una farfalla o se è una farfalla che sogna di essere Tzu». Che prove avete, voi che leggete queste righe, di avere veramente un libro fra le mani e di non essere vittime di una perfetta illusione. In realtà potreste essere un cervello in una vasca piena di liquido nutriente e collegato a un computer che stimola i vostri neuroni in modo da farvi credere che state leggendo un libro. L'associazione con il film *Matrix* è scontata, ma non è scontata, perché nota solo agli addetti ai lavori quella con un saggio del filosofo americano Putnam il quale in *Cervelli in una vasca*, appunto, vi sfida a dimostrare che non siete cervelli in una vasca (non è facile vincere tale sfida, ma Putnam ha scritto il suo saggio perché pensa di averla vinta)⁵. Un problema strettamente connesso al precedente è quello che aveva in mente Descartes (Cartesio) quando si chiedeva (più o meno): «cosa mi garantisce che quei cappelli e quei mantelli che vedo dalla mia finestra passare giù per strada coprono degli uomini veri e non degli automi?». Descartes pensava di possedere un criterio per rispondere a tale domanda: gli uomini agiscono e parlano in modo creativo, gli automi in modo ripetitivo.

Secondo Descartes il pensiero umano esibisce una caratteristica che non può essere ricondotta a una meccanica «disposizione d'organi»: questa caratteristica è l'originalità, o creatività, che si manifesta soprattutto nel linguaggio. Poiché gli uomini sono in grado di dire e capire frasi che non hanno mai detto o udito in precedenza e poiché, invece, un meccanismo che parlasse lo potrebbe fare solo ripetendo espressioni che ha già interiorizzato, allora bisogna concludere che nell'uomo è presente un principio immateriale con caratteristiche di originalità e creatività. Questo principio è per Descartes l'anima.

Mi ero qui fermato a far vedere in modo particolare che, se ci fossero macchine aventi organi e figura di scimmia o di altro animale privo di ragione, noi non avremmo nessun mezzo per riconoscere la differenza, mentre, se ve ne fossero che somigliassero al nostro corpo e imitassero le nostre azioni quanto meglio fosse possibile, noi avremmo pur sempre dei mezzi certissimi per riconoscere che esse non sono affatto per ciò dei veri uomini. Il primo è che non potrebbero mai valersi di parole o di altri segni, componendoli come noi facciamo per esprimere agli altri i nostri

pensieri: poiché si può ben immaginare una macchina che profferisca delle parole, e anzi ne profferisca alcune riguardanti azioni corporali che producano qualche alterazione nei suoi organi, come domandare qualcosa se toccata in una parte, o gridare che le si fa male se toccata in altra parte, e simili cose; ma non già che essa disponga le parole diversamente per rispondere a tono a tutto quello che uno può dirle, come, invece saprebbe fare anche l'uomo più idiota. Il secondo mezzo è che, anche se facessero alcune cose ugualmente bene e anzi meglio di noi, esse inevitabilmente sbaglierebbero in alcune altre, e si scoprirebbe così che non agiscono per conoscenza, ma solo per una disposizione dei loro organi. Soltanto la ragione, infatti, è uno strumento universale, che può servire in ogni specie di circostanze e tali organi, invece, hanno bisogno di una particolare disposizione per ogni azione particolare: sì che è come impossibile che ce ne siano tanti e così diversi in una macchina da farla agire in ogni occasione nel modo che agiamo noi con la nostra ragione [6, pp. 40-41]

Ciò significa che i robot e i computer non possono pensare. Ma allora le storie di fantascienza che hanno come protagonisti robot e computer sono solo fantastiche. Non esattamente: come dicevo all'inizio occorre aspettare la seconda metà del XX secolo perché l'ipotesi di robot e computer pensanti potesse essere presa in seria considerazione.

3. AI, *Blade Runner* e 2001 *Odissea nello spazio*

Il film *A.I. Intelligenza artificiale* (2001) di S. Spielberg realizza un antico progetto di S. Kubrick (il regista di *2001 Odissea nello spazio*) a sua volta ispirato a tre racconti dello scrittore di fantascienza B. W. Aldiss⁶. La storia può essere considerata una versione fantascientifica di Pinocchio (e le citazioni del Pinocchio di Walt Disney sono esplicite). In un futuro indeterminato, nel quale la tecnologia è avanzatissima ma le risorse ormai limitate, il figlio dei coniugi Swinton, Martin, si ammala di una male incurabile e viene ibernato in attesa che la scienza trovi un rimedio. I genitori Henry e Monica vengono selezionati dal professor Hobby per sperimentare un nuovo modello di robot bambino in grado non più soltanto di simulare emozioni ma di provarle veramente e quindi in grado di amare i suoi «genitori» adottivi. Il robot bambino, David, fa l'ingresso nella vita della coppia Swinton dove, in breve tempo, riesce a vincere l'iniziale diffidenza di Monica. Tanto sono indistinguibili l'aspetto e i comportamenti di David da quelli di un bambino vero che Monica gli si affeziona. Da parte sua, David è programmato per amare la madre adottiva.

Tra Monica e David si stabilisce dunque un vero e proprio legame affettivo. Un giorno, però, la medicina trova la cura per la malattia di Martin che guarisce e fa ritorno in famiglia. David capisce che il rapporto fra Monica e Martin è speciale e comincia a pensare che se anche lui diventasse un bambino vero potrebbe avere lo stesso amore che Monica nutre per Martin. Una serie di equivoci e di malfunzionamenti costringono Henry e Monica a disfarsi di David. David comincia così la sua ricerca di qualcuno

che lo possa trasformare in un bambino vero, comincia cioè la sua ricerca della Fata Turchina. Ne trova una statua in un parco giochi sommerso dal mare e, credendola vera, resta in attesa che lo trasformi in bambino. Passano duemila anni, una glaciazione ha distrutto la civiltà, David è entrato in stato di sospensione, l'umanità si è estinta e al suo posto è subentrata una razza di robot costruiti da generazioni di altri robot. I robot ritrovano David e lo «riportano in vita». David riesce ad ottenere da loro un clone di Monica che però può vivere solo un giorno. Il film si chiude sull'immagine di David che tiene la mano a Monica mentre si addormenta per sempre.

Blade Runner di R. Scott (1982) è ispirato al romanzo di P. K. Dick *Gli androidi sognano pecore elettriche?* (1968). Il film, e il romanzo, raccontano la storia di androidi (i replicanti, fabbricati con materiali biologici di origine sintetica) esteriormente indistinguibili dagli esseri umani. Poiché alcuni di questi androidi si comportano da criminali sono state istituite delle unità di polizia (dette *Blade Runner* appunto) per individuarli ed eliminarli. Per riconoscere un replicante da un essere umano la polizia sottopone i sospetti a un test, il test Voikampf. Si tratta di una sorta di macchina della verità che permette di analizzare nei dettagli le reazioni fisiche dei sospetti a opportune domande. Evidentemente le reazioni dei replicanti risulteranno più «fredde» di quelle degli esseri umani.

All'origine di molti capolavori, il genio di Stanley Kubrik si è cimentato anche con il genere fantascientifico. Il celeberrimo *2001 Odissea nello spazio* (1968) è tratto da un racconto di Arthur C. Clarke, *La sentinella*, 1953⁷. Il protagonista del film di Kubrik, il calcolatore Hal 9000, è di aspetto completamente diverso da quello di David e dei replicanti. Fosse per il suo aspetto, Hal non ingannerebbe nessuno: non un ha corpo simile a quello di un essere umano, è semplicemente un calcolatore di grandi dimensioni con i suoi bravi pulsanti e spie luminose bene in evidenza. L'interfaccia fra uomo e calcolatore è l'obiettivo di una sorta di telecamera, una lente, l'«occhio» attraverso il quale Hal guarda l'ambiente in cui si trova che è un'astronave in viaggio verso Giove. Eppure Hal riesce non solo a svolgere compiti logico-matematici, come ogni bravo calcolatore, mantenendo per esempio in rotta l'astronave ma sembra provare anche emozioni, come l'orgoglio e la paura. L'orgoglio della perfezione delle sue prestazioni viene infatti un giorno ferito da una previsione sbagliata circa l'avaria del componente di un'antenna dell'astronave. Preoccupati dalla possibilità che Hal possa commettere errori più gravi gli astronauti decidono di disattivarlo. All'orgoglio subentra in Hal la paura e per impedire di essere disattivato uccide gli astronauti. Tutti tranne uno, David (lo stesso nome del protagonista di *AI*) il quale disattiva Hal e continua la missione.

4. Quali esseri hanno una mente?

Negli anni Cinquanta del Novecento il matematico Alan Turing – uno dei fondatori della moderna IA – si propose di affrontare in modo nuovo la questione che dà il titolo al paragrafo e scriveva:

Mi propongo di considerare la domanda: «possono pensare le macchine?» [...]

La nuova forma del problema può essere descritta nei termini di un gioco, che chiameremo il «gioco dell'imitazione». Questo viene giocato da tre persone, un uomo (A), una donna (B) e l'interrogante (C), che può essere dell'uno o dell'altro sesso. L'interrogante viene chiuso in una stanza, separato dagli altri due. Scopo del gioco per l'interrogante è quello di determinare quale delle altre due persone sia l'uomo e quale la donna. Egli le conosce con le etichette X e Y, e alla fine del gioco darà la soluzione «X è A e Y è B» o la soluzione «X è B e Y è A». L'interrogante può far domande di questo tipo ad A e B: «vuol dirmi X, per favore, la lunghezza dei propri capelli?» Ora supponiamo che X sia in effetti A, quindi A deve rispondere. Scopo di A nel gioco è quello di ingannare C e far sì che fornisca una identificazione errata [...] Le risposte, in modo che il tono della voce non possa aiutare l'interrogante, dovrebbero essere scritte, o, meglio ancora, battute a macchina. [...]

Poniamo ora la domanda: «che cosa accadrà se una macchina prenderà il posto di A nel gioco?» L'interrogante darà una risposta errata altrettanto spesso di quando il gioco viene giocato tra un uomo e una donna? Queste domande sostituiscono quella originale: «possono pensare le macchine?». Il metodo delle domande e risposte sembra essere adatto per introdurre nell'esame quasi ogni campo della conoscenza umana che desideriamo. Non desideriamo penalizzare la macchina per la sua incapacità di brillare in un concorso di bellezza, né penalizzare un uomo perché perde una corsa contro un aeroplano. Le condizioni del nostro gioco rendono irrilevanti queste incapacità [19, p. 116]

I replicanti di *Blade Runner* sospetti di attività criminali vengono sottoposti ad un test simile a quello di Turing al quale, anzi, il test di Voitekampf è ispirato e del quale rappresenta la versione letteraria e cinematografica. Il test suppone che un essere che si comporti (dove il «comportamento» comprende anche le risposte verbali alle domande) in modo indistinguibile da un essere umano *sia* un essere umano⁸. Sicuramente tale supposizione è condivisa dai poliziotti dell'unità *Blade Runner* i quali non dubitano della validità del test. Diversa la posizione degli scienziati che hanno progettato Hal 9000. Uno di costoro, intervistato da un giornalista il quale chiede se Hal abbia emozioni, afferma che «si comporta come se le avesse, ma se le abbia davvero o no nessuno può dirlo con certezza». Nella prima scena di *AI* il professor Hobby espone ai collaboratori il progetto di costruire robot che provino emozioni e, per sottolinearne la novità, rivolge a un robot di vecchia generazione, identico nell'aspetto e nei comportamenti ad un essere umano ma incapace di provare emozioni, la domanda «Cos'è l'amore?» ottenendo la risposta: «l'amore è sgranare leggermente gli occhi, accelerare la respirazione, scaldare la pelle...».

In effetti superare il test non costituisce una prova decisiva del possedere pensieri ed emozioni poiché il test valuta comportamenti che potrebbero essere effettuati anche da una macchina priva di mente: tale macchina *simulerebbe* il provare, per esempio, dolore perché farebbe e direbbe tutto ciò che un essere umano fa e dice quando prova dolore senza tuttavia provare realmente dolore. Ciò che sfugge al test è proprio ciò che per gli esseri umani conta: l'aver pensieri e sensazioni, l'aver una vita interiore che,

insomma, equivale all'averne una vita *tout court*. L'amore non è solo «sgranare leggermente gli occhi, accelerare la respirazione, scaldare la pelle» ma provare un sentimento. In generale avere stati mentali – e, d'ora in poi, intendo per stati mentali tutto ciò che avviene nella nostra mente: pensieri, ricordi, sensazioni, sentimenti, consapevolezza ecc. – non è ridicibile ai comportamenti che sono ad essi associati. Qualcuno, o qualcosa, potrebbe comportarsi come se pensasse o amasse ma in realtà non avere affatto una vita mentale. Questa, riteniamo, sia la situazione di una macchina (robot o calcolatore) o di uno zombi. È dunque sempre possibile dire dei robot, dei computer e magari degli alieni che simulano il comportamento umano: diciamo che essi si comportano *come se* avessero una mente, ma in realtà non ce l'hanno. Ma perché non diciamo lo stesso anche degli esseri umani? Forse perché sappiamo che nel loro cranio c'è qualcosa di speciale, responsabile della vita mentale: il cervello?

Se si considera il possesso di un sistema nervoso un elemento indispensabile per il possesso di una mente è naturale rivolgersi alle neuroscienze nella speranza di ottenere lumi per capire come «l'acqua del cervello si cambi nel vino della mente» [12, p. 173]. È questa la speranza di cui si è nutrito e si nutre quella forma di materialismo contemporaneo che va sotto il nome di teoria dell'identità⁹. Si tratta di una teoria nata alla fine degli anni Cinquanta del secolo scorso la quale sostiene che gli stati mentali sono identici a stati cerebrali. Questa forma di materialismo si differenzia da quella classica – rappresentata nella filosofia moderna per esempio da Hobbes – perché non sostiene una sua propria ontologia ma, piuttosto, affida alla ricerca delle scienze naturali il compito di pronunciarsi su quali siano i componenti ultimi della realtà. L'identità fra stati mentali e stati cerebrali è ritenuta una ipotesi la verifica o la falsificazione della quale deve essere affidata alla ricerca empirica, in particolare alle neuroscienze. Quasi tutto il lavoro teorico dei sostenitori dell'identità è stato infatti indirizzato a difendere la loro posizione da obiezioni a priori. Si tratta di obiezioni riconducibili, più o meno direttamente, alla cosiddetta legge di Leibniz secondo la quale se due entità sono identiche allora avranno le stesse proprietà. Descartes, per esempio, sosteneva che la mente è distinta e diversa dal corpo perché mente e corpo hanno proprietà diverse. I corpi, per esempio, sono estesi ma le menti no; la mente è invece in grado di pensare mentre il corpo no. I teorici dell'identità hanno cercato di mostrare che tali argomentazioni sono, in realtà, riconducibili a petizioni di principio: si assume la verità di ciò che deve essere dimostrato. Se si assume già che la mente non sia estesa e che la materia non possa pensare, è banale poi concludere che la mente e la materia sono distinte.

La sensibilità mostrata dalla teoria dell'identità per i risultati della ricerca scientifica potrebbe indurci a pensare che la teoria abbia riscosso un ampio consenso: ma ciò che è avvenuto è esattamente il contrario. I sostenitori della teoria sono rimasti un manipolo esposto al fuoco incrociato proveniente da più parti. L'esame di alcune di tali obiezioni ci permetterà di entrare un po' più in dettaglio nel problema mente corpo.

Nel film *Viaggio allucinante*¹⁰, si immagina che sia possibile ridurre a dimensioni

microscopiche una sorta di sottomarino e di introdurlo, mediante iniezione endovenosa, nel circolo sanguigno di un essere umano. L'equipaggio del sottomarino, giunto al cervello, osserva ciò che ha osservato in ogni altra parte del corpo: tessuti, liquidi organici e cellule di ogni tipo. Ora, mentre l'osservazione di tali oggetti è ciò che è desiderabile osservare per spiegare come funziona, per esempio, il sistema immunitario o il cuore, non è di nessun aiuto per tentare di spiegare l'origine degli stati mentali dal cervello. Per quanto si osservi da vicino il cervello non si osserverà nessun pensiero. Ciò significa che i pensieri non hanno a che fare con il cervello? Nei secoli XVII e XVIII Descartes e Leibniz avevano risposto affermativamente a tale domanda. Leibniz, in particolare, aveva proprio immaginato una sorta di *Viaggio allucinante*. «Immaginiamo - scrive Leibniz nella *Monadologia* del 1714 - una macchina in grado di pensare, sentire, percepire; ed immaginiamo che sia così grande da poterci entrare a visitarla, come in un mulino. Visitandone l'interno non si troveranno altro che pezzi di materia che si spingono a vicenda, ma non si troverà nulla che possa spiegare la percezione»¹¹. Una replica abbastanza ovvia, che difatti è stata avanzata, consisterebbe nel far notare che neppure la liquidità dell'acqua sarebbe percepibile se navigassimo sott'acqua con un sottomarino più piccolo delle molecole di idrogeno e di ossigeno: eppure l'acqua è liquida. Ma si potrebbe anche far notare che il pensiero non è qualcosa che si *osserva* in nessun senso ordinario del verbo *osservare*. Ora il problema è: questa considerazione vale pro o contro il materialismo? Dipende - e su questo il dibattito è in corso -. Se la «non-osservabilità-ordinaria» è una proprietà oggettiva del mentale allora, considerata anche l'«osservabilità-ordinaria» una proprietà oggettiva dei corpi, si può ricorrere alla legge di Leibniz per sostenere il dualismo. Se invece consideriamo le due modalità dell'osservabilità due diverse modalità di conoscere una stessa cosa, allora non se ne può ricavare la diversità fra stati mentali e stati del sistema nervoso più di quanto si possa ricavare la diversità fra Kubrik e il regista di *2001 Odissea nello spazio* perché io l'ho conosciuto personalmente e voi solo attraverso i suoi film.

Il possedere un cervello sembra essere così importante che siamo disposti a dire che un essere umano ha una mente anche se si comporta in modo abnorme: caso limite, non abbiamo tanta difficoltà ad attribuire una mente a malati completamente paralizzati affetti da distrofia amiotrofica laterale quanta ne abbiamo ad attribuirgli ad un computer. Il problema etico dell'eutanasia, anzi, nasce perché sospettiamo che gli esseri umani possano continuare ad avere una mente anche in quei casi estremi in cui non la manifestano più. Siamo così passati dal considerare il comportamento e l'aspetto esteriore di un essere come indizi (non è il caso di usare la più impegnativa parola «prova») del possesso di una mente all'esame della sua costituzione interna.

Chiamati a decidere se un essere possiede o no una mente sembra dunque che utilizziamo un criterio di somiglianza e che ragioniamo così: «io ho una mente, quegli esseri somigliano a me, dunque quegli esseri hanno una mente». Ma in che senso intendiamo «somigliano», nell'aspetto esteriore e nel comportamento o nella costituzione interna? Il criterio della somiglianza che impieghiamo spontaneamente

per decidere se un essere ha una mente sembra infatti avere due componenti: un essere può essere simile a un altro nei comportamenti (reazione agli stimoli, discorsi, abilità, capacità di agire ecc.) e nella costituzione interna (materiale, organi, apparati ecc.). Evidentemente tutti gli esseri umani soddisfano entrambe le componenti, ma che dire di robot e alieni? Se ci pensiamo bene, tuttavia, il problema si pone già per esseri che non sono prodotti della fantasia. Anche per organismi che hanno una costituzione interna simile alla nostra, e ciò in sostanza significa che possiedono un sistema nervoso, non è così semplice stabilire se hanno o no stati mentali. Dov'è, nel regno animale, che cade la demarcazione fra organismi che possiedono stati mentali e organismi che non li possiedono? Se tutti hanno sistemi nervosi allora la demarcazione ha a che vedere con la complessità di quei sistemi?

Se cominciamo a considerare la questione anche da un punto di vista biologico, i problemi si moltiplicano: la domanda non è più solo «come può la coscienza sorgere da una molliccia materia grigia» ma anche «a cosa serve la coscienza?» Infatti la maggior parte degli esseri viventi non la possiede: milioni (letteralmente *milioni*: solo le specie attualmente note di insetti superano il milione¹²) di specie animali vivono benissimo senza e non sembrano avere nessuna intenzione di soccombere alle specie che la possiedono (ma poi, quante oltre alla specie umana?). Dov'è dunque il vantaggio biologico, selettivo, evolutivo dell'avere la coscienza? La questione non è banale poiché ci sono state e ci sono posizioni (dette epifenomeniste) che negano che la coscienza abbia una funzione anche per gli organismi che ne sono dotati. Secondo tale posizione la coscienza è un modo con cui veniamo a sapere che cosa accade nella nostra mente/cervello, ma non ha nessuna efficacia causale sui nostri comportamenti. Sarebbe come il guardare in uno schermo per computer: tramite esso vediamo ciò che succede nel computer ma non possiamo influenzarlo (quando lo facciamo, usiamo la tastiera). Si è sostenuto che l'inefficacia causale della coscienza non è in contrasto con la teoria dell'evoluzione: la situazione sarebbe analoga a quella della pelliccia degli orsi polari. La pelliccia degli orsi polari è calda e svolge egregiamente la sua funzione di proteggere l'animale dal freddo, tuttavia è anche pesante e ingombrante: ma se volete una pelliccia calda e confortevole dovete anche tenervi una pelliccia pesante e ingombrante. Penso però che si debba riflettere sul fatto che gli esseri umani privi di coscienza, per ragioni patologiche come per esempio il coma, non riescono a fare ciò che fanno gli esseri umani normali. Questo renderebbe insostenibile qualunque confronto fra esseri umani e animali privi di coscienza poiché negli organismi che ce l'hanno essa avrebbe efficacia causale.¹³

La presenza di un sistema nervoso in un organismo potrebbe accompagnarsi a comportamenti così diversi dai nostri che potremmo essere in imbarazzo nell'attribuzione di stati mentali a quell'organismo. Un filastrocca per bambini recita:

*C'era una volta un ranocchietto di stagno,
giù nell'acqua scura scura
tutta notte stava a bagno;*

*quando usciva all'aria pura
d'ogni cosa aveva paura*

Evidentemente l'aver paura è uno stato mentale e quindi il ranocchio della filastrocca ha una mente (foss'anche di un solo stato): ma quanto diverso il suo comportamento dal nostro nel preferire l'acqua scura e fredda alla luce e all'aria pura! E quanto opposte alle nostre le sue opinioni in merito ai posti salubri! Sappiamo però dall'anatomia che il ranocchio possiede un sistema nervoso centrale, e questo ci trattiene dal negargli decisamente il possesso di stati mentali (non importa quanto semplici): infatti non neghiamo che abbia stati mentali ma diciamo che ha stati mentali diversi dai nostri. In generale attribuiamo stati mentali agli animali più per il fatto che sappiamo che possiedono un sistema nervoso che per le somiglianze esteriori o del comportamento. E ciò sembra corretto perché non possiamo pretendere, peccando di antropocentrismo, che solo esseri del tutto identici a noi abbiano stati mentali. Il che equivarrebbe poi a dire che in tutto l'universo solo noi umani abbiamo una mente.

L'aver una mente non è una proprietà isolabile da altre proprietà dell'organismo e dell'ambiente in cui vive. Per limitarsi al solo il caso della percezione esistono animali che impiegano modalità percettive del tutto sconosciute alla specie umana. Per esempio, alcuni serpenti delle famiglie Boidi e Crotalidi sono dotati di termorecettori che consentono loro di individuare a distanza la possibile preda dal calore del corpo. Essi avranno, dunque, una sorta di «rappresentazione termica del mondo» del tutto inimmaginabile a noi che ne abbiamo una rappresentazione prevalentemente «visiva». Ma, come è facile aspettarsi, il caso più famoso, e anche più citato anche nella letteratura filosofica sulla mente, è quello del sistema sonar dei pipistrelli. I dati forniti dalle ricerche in quest'ambito rendono difficile sostenere che gli stati mentali degli animali siano simili ai nostri. La considerazione è rilevante se si pensa che non tutte le teorie della mente considerano il sistema nervoso, e in generale la prospettiva biologica, elemento caratteristico per il sorgere del mentale. Queste teorie, dette funzionaliste, ritengono che la mente sia una sorta di programma per computer e che quindi l'hardware – il sistema nervoso - su cui il programma gira sia del tutto secondario.

5. Pensare e calcolare

Gli stati mentali ai quali ci siamo fin qui riferiti hanno una caratteristica in comune, quella di avere una qualità. Vedere un colore, avere paura, sentire freddo, provare un dolore, amare ecc. fanno tutti un certo effetto a chi li possiede. Ma vi sono stati mentali che non si distinguono per avere una particolare qualità: per esempio il cercare di risolvere un'equazione di secondo grado non fa nessun effetto. Almeno non in modo costitutivo. Può, sì, essere associato a stati qualitativi: posso per esempio provare soddisfazione nel pervenire alla soluzione, o ansia nel dover consegnare il compito non risolto, ma si tratta di sensazioni associate al calcolare: il calcolare come tale non fa nessun effetto. Eppure, e questo è ciò che ci interessa, calcolare è sicuramente uno stato mentale, è una forma di pensiero. Anzi è una delle forme più alte di pensiero. Eminentemente

filosofi hanno considerato il calcolare come il modello stesso del pensare. È superfluo dire che senza quest'idea (nata con Hobbes e Leibniz in età moderna e sviluppata nel corso XX secolo con strumenti logico matematici più raffinati) non avremmo i computer. Infatti i computer – lo dice il nome – computano, contano, calcolano. Ma se calcolare è una forma pensiero allora ne dovrebbe seguire che i computer pensano. Questo è il ragionamento che sta dietro alle storie di fantascienza che hanno per protagoniste computer e robot. Forse Hal 9000 pensa; ma se «lui» pensa perché calcola allora pensava anche la calcolatrice di Pascal, una macchinetta a manovella composta da ingranaggi che assomigliava più a un macinino per caffè che a un computer?

C'è una teoria contemporanea della mente che fa dell'idea che pensare sia calcolare la sua forza: è il funzionalismo o teoria computazionale¹⁴ della mente cui ho accennato sopra. Per il funzionalismo gli stati mentali non sono stati del sistema nervoso, ma stati funzionali. Uno stato funzionale è uno stato che può essere individuato grazie alle sue relazioni con altri stati dell'organismo, cioè con i dati sensoriali che lo causano (input), con i comportamenti che, a sua volta, esso causa (output) e con altri stati mentali (credenze, opinioni, ricordi eccetera). La rete di sensazioni, comportamenti e stati mentali che definiscono un certo stato mentale è caratteristica di tale stato. Se, per esempio Giovanni si trova nello stato «avere mal di denti» allora Giovanni si trova in uno stato tale che 1. è causato da: avere mangiato troppi dolci e avere una carie; 2. è causa di: lamenti, premersi una mano sulla guancia, prendere un analgesico, telefonare al dentista; 3. è associato a: credere che un analgesico calmerà il dolore, credere che il dentista abbia una laurea in medicina ecc. In generale uno stato mentale x è definito in funzione di altri elementi v, w, y, z per cui $x = f(v, w, y, z)$.

Considerati come stati funzionali gli stati mentali possono essere concepiti come programmi per computer, dal momento che gli stati funzionali sono concepibili come programmi per computer. Definiamo, semplicemente, un programma per computer come una lista di istruzioni tale che, a partire da certi input e certi stati interni produce certi output¹⁵. Ora, quello che è decisivo per la filosofia della mente, è il fatto che i programmi sono nozioni astratte realizzabili fisicamente in molti modi diversi. Il materiale di cui è costituito un computer è indifferente: quello che conta è il programma che può essere implementato in supporti materiali diversi. Per il funzionalismo la mente sta al software come il cervello sta all'hardware: uno stesso software può essere realizzato in macchine materialmente diverse. Il cervello umano è solo uno dei possibili supporti materiali in cui può essere implementato un programma che descrive i nostri stati mentali. Se è così, allora, gli stati mentali non sono stati del sistema nervoso più di quanto non possano essere stati di un sistema elettronico, idraulico, meccanico...¹⁶ Questa tesi, caratteristica del funzionalismo, è nota come tesi della realizzazione molteplice degli stati mentali. Si capisce dunque perché il funzionalista sia più disposto di altri ad assegnare stati mentali a robot, a computer e ad alieni. Solo adottando tale tesi è possibile attribuire stati mentali a esseri fisicamente diversi da noi (animali, alieni, artefatti).

Il principale problema del funzionalismo nasce dal fatto che la descrizione del mal di denti di Giovanni (per riprendere l'esempio sopra riportato) sarebbe corretta anche nel caso in cui Giovanni fingesse di avere mal di denti oppure soffrisse, stoicamente, senza manifestarlo. Quello che sfugge alla descrizione funzionale è la «dolorosità» del mal di denti, cioè il fatto che il mal di denti «faccia male». Questo aspetto, qualitativo, è invece proprio ciò che conta nello stato mentale di Giovanni: l'aver un mal di denti è esattamente il *sentire* un dolore. In generale gli stati qualitativi (*qualia*) costituiscono l'ostacolo più serio in cui il funzionalismo si è imbattuto. Gli esperimenti mentali elaborati intorno a questo problema sono numerosi, ne ricorderò tre proprio perché sono buoni esempi di esperimenti mentali

Siano A e B due organismi che percepiscono lo spettro dei colori invertito: ad A i pomodori maturi e gli automezzi dei pompieri appaiono dello stesso colore in cui a B appare il colore dell'erba e delle foglie. D'altra parte ad A il colore dell'erba appare dello stesso colore in cui a B appaiono gli automezzi dei pompieri. Come potremmo accertare che A vede «verdi» i camion dei pompieri? Interrogato risponderà che il colore di tali oggetti è «rosso» poiché gli è stato insegnato che il nome del colore degli automezzi dei pompieri è «rosso», nonostante il fatto che gli appaiano verdi. Insomma A e B avranno le stesse credenze e si comporteranno nello stesso modo di fronte a oggetti che vedono di colori diversi. Dunque le descrizioni funzionali di A e B saranno identiche nonostante il fatto che A e B abbiano stati qualitativamente diversi. La descrizione funzionale non riesce pertanto a cogliere l'aspetto più importante della visione: la qualità del colore percepito.

Nel secondo esperimento mentale si tenta di mostrare come la descrizione funzionale di un organismo che abbia un certo stato qualitativo sia identica a quella di un organismo analogo che però manchi di tale stato. Ad esempio N. Block immagina 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 Giovanni. Ogni cinese realizzerà un passo di tale programma e quindi la descrizione funzionale della Cina sarà identica a quella di Giovanni quando ha il mal di denti. Ma la Cina non proverà un mal di denti. Anche in questo caso, dunque, la descrizione funzionale fallisce nel cogliere l'aspetto caratteristico di quello stato mentale: la qualità del dolore.

Come ho detto, argomenti simili a quelli precedenti sono, nella letteratura specifica, numerosi (una versione del primo è d'altronde già reperibile in Locke). Ma la critica più diffusa contro il funzionalismo e, in genere, contro ogni teoria computazionale della mente, è quella avanzata da J. Searle. Egli ha rovesciato il test di Turing cercando di mostrare che i computer non possono pensare e che, siccome gli esseri umani pensano, le teorie computazionali sono false in quanto considerano la mente umana come un computer. Allo scopo di decidere se una macchina abbia stati mentali non è necessario né sufficiente osservare i comportamenti della macchina o analizzarne gli stati interni: più semplicemente esaminiamo ciò che avviene negli esseri umani quando questi si comportano come una macchina. In altre parole per sapere se i computer pensano

dobbiamo metterci nei panni del computer. È questo l'argomento della stanza cinese e suona come segue:

Supponiamo che io sia chiuso dentro una stanza e che mi si dia una serie di fogli scritti in cinese. Supponiamo inoltre (come infatti è il caso mio) che non conosca il cinese, né scritto né parlato e che, anzi, non sia nemmeno in grado di poter riconoscere uno scritto cinese in quanto tale, distinguendolo magari dal giapponese o da scarabocchi senza senso. Per me la scrittura cinese è proprio come tanti scarabocchi senza senso. Ora supponiamo ancora che mi sia dia un secondo pacco di fogli, sempre scritto in cinese, unitamente ad una serie di regole per mettere in relazione il secondo plico con il primo. Le regole sono nella mia lingua e io capisco queste regole come qualunque altro parlante la mia stessa lingua. Esse mi rendono possibile mettere in relazione una serie di simboli con una serie di altri simboli esclusivamente in virtù della loro forma. Ora supponiamo che mi si dia una terza serie di simboli cinesi con le relative istruzioni, sempre nella mia lingua, che mi rendano possibile correlare elementi di questo terzo pacco con i primi due, e che queste regole mi istruiscano su come riprodurre certi simboli cinesi con certi tipi di forme datemi nel terzo plico. A mia insaputa, le persone che mi danno tutti questi simboli chiamano il primo pacco di fogli «uno scritto», chiamano il secondo «una storia» e il terzo «quesiti». Inoltre chiamano i simboli che rendo loro in risposta al terzo plico «risposte ai quesiti», e la serie di regole nella mia lingua che mi hanno dato la chiamano «il programma». [...] Supponiamo anche che io diventi così bravo nel seguire le istruzioni per manipolare simboli cinesi e che i programmatori diventino così bravi nello scrivere i programmi che dal punto di vista esterno – cioè dal punto di vista di qualcuno al di fuori della stanza nella quale sono chiuso – le mie risposte alle domande assolutamente non si distinguano da quelle di cinesi madrelingua. Nessuno che guardi bene alle mie risposte può dire che io non parli una parola di cinese [...] ma produco le risposte con il manipolare simboli formali non interpretati. Per quanto riguarda il cinese, mi comporto semplicemente come un computer: eseguo operazioni calcolabili su elementi formalmente specificati. Per il caso del cinese, io sono semplicemente una istanziazione di un programma per computer. [18, pp.48-49]

Il punto essenziale della posizione di Searle è che se, come il suo argomento vorrebbe dimostrare¹⁷, la mente non può derivare dal software allora deve derivare dall'hardware, cioè da qualche proprietà fisico chimica del cervello.

In parte per i suoi problemi interni, in parte per il progresso delle ricerche sull'intelligenza artificiale, il funzionalismo ha cominciato a lasciare spazio a un'altra prospettiva che costituisce a tutt'oggi una ricerca in corso: il connessionismo o teoria delle reti neurali. Si tratta ancora di una teoria computazionale ma temperata da una tesi tipicamente materialistica, quella, cioè, di prendere in considerazione i risultati delle ricerche delle neuroscienze. Anche il connessionismo ritiene, infatti, con il funzionalismo, che gli stati mentali siano stati di un programma, ma, con la teoria dell'identità, tiene presenti le caratteristiche fisico-biologiche del cervello umano. Ad esempio il fatto che il cervello umano è, rispetto ai computer, lento nell'elaborare le

informazioni, oppure il fatto che il cervello è eccellente nel riconoscere le forme ma, rispetto a un computer, ha scarse abilità di calcolo. Ma la caratteristica più saliente delle reti neurali risiede nella loro capacità di apprendere: come il cervello, esse non hanno bisogno di essere preventivamente programmate. Se questa nuova prospettiva si affermerà definitivamente allora la contrapposizione fra sostenitori e denigratori dell'idea che la mente è un computer potrà essere superata: la risposta non sarà né sì né no, ma un po' sì e un po' no.

5. Zombi

In quanto abbiamo veduto sopra si è fatto largo uso della fantasia, degli esperimenti mentali e della fantascienza. Spero che sia risultata evidente l'utilità di tali procedure, le quali sovente ci permettono di chiarire i problemi scorgendone aspetti nascosti. In questo senso esse hanno un autentico valore euristico. Tuttavia vorrei – saggiamente – riequilibrare le cose e concludere con l'avvertenza che la fantasia può portarci anche là dove non dovremmo andare, se vogliamo appunto fare scienza e non solo fantascienza. Per far questo mi servirò degli zombi. D'altra parte, non è possibile intervenire su cinema e fantascienza senza menzionare gli zombi.

Anche in questo caso siamo a che fare con un esperimento mentale che intende mostrare che gli stati mentali non sono stati cerebrali. Gli zombi della filosofia non assomigliano a quelli della fantascienza o dei film dell'orrore: nei film è infatti facile capire se qualcuno è uno zombi: si muove in modo impacciato, ha un colorito verdastro e perde i pezzi. Gli zombi della filosofia sono completamente indistinguibili dagli esseri umani, non solo come aspetto e come comportamento, ma se ne fate l'autopsia trovate che all'interno sono identici a noi. L'unica differenza sta nel fatto che sono privi di stati mentali e di coscienza. In casa non c'è nessuno. Ora l'argomento è: se sono concepibili gli zombi come appena descritti, cioè esseri fisicamente identici a noi ma privi di coscienza, allora la coscienza non è implicata da nessun fatto fisico. Scrive l'autore di tale argomento:

Consideriamo il mio gemello zombi. Questa creatura è identica a me molecola per molecola e per tutte le proprietà di basso livello postulate da una fisica completa, ma lui è totalmente privo di esperienza conscia [...] Egli è identico a me fisicamente, e possiamo anche supporre che sia immerso in un ambiente identico. Certamente è identico a me funzionalmente [...] e dà luogo agli stessi comportamenti indistinguibili. È psicologicamente identico a me [...] è vigile, capace di riferire il contenuto dei propri stati interni, capace di concentrare l'attenzione sui vari luoghi e così via. Semplicemente nessuno di questi processi è accompagnato da una qualsivoglia esperienza conscia autentica. Non c'è un sentire in senso fenomenico. Non si prova nulla a essere uno zombi [3, p. 97].

Sicuramente un tale essere è *immaginabile* e ciò, continua Chalmers, è sufficiente per costruire un argomento antiriduzionista:

1. se gli zombie sono immaginabili,

2. allora sono possibili;
3. se sono possibili
4. allora possono esistere;
5. se possono esistere allora il fisicalismo è falso.

Anche la discussione su tale argomento è accesa e ha prodotto molta letteratura nella filosofia della mente contemporanea, ma l'impressione è che tale dibattito sia divenuto autoreferenziale. Da un punto di vista meramente logico tutti i contendenti possono reperire argomenti per continuare a sostenere le loro posizioni. Concetti come quelli di possibilità e necessità metafisica, logica e reale sono complessi e intrattengono relazioni complicate sulle quali non c'è unanimità. Non è ancora sufficientemente chiaro quali siano i confini e i rapporti fra ciò che è immaginabile, ciò che è concepibile, ciò che è logicamente possibile e ciò che è realmente possibile. C'è tuttavia, a mio avviso, un modo relativamente semplice di non smarrirsi nel *fog of war* sollevato dalla discussione ed è quello di riprendere l'ispirazione di fondo del fisicalismo, cioè di rifarsi ad un approccio naturalistico. Immaginiamo un'automobile-zombi: essa è esattamente come un'auto normale, atomo per atomo, e tuttavia – direbbe Chalmers - mentre l'auto normale si mette in moto l'auto-zombi non parte. Si possono immaginare due auto identiche atomo per atomo una delle quali non parta? Bhè, potremmo appunto scriverci sopra un racconto di fantascienza intitolato *Lo strano caso dell'auto-zombi* ma la conclusione non può essere che una: alla fine un meccanico-eroe scopre un guasto che nessun era riuscito a scoprire. Ora, cosa c'è di diverso fra il caso dell'auto-zombi e quello dell'uomo-zombi? A mio avviso niente, altrimenti non ho capito l'argomento. La plausibilità (apparente) dell'argomento degli zombi deriva da secoli di dualismo: viviamo in una cultura che da secoli crede che la mente sia diversa dal corpo e questo ci ha familiarizzato con l'idea che la mente e il corpo possano esistere separatamente. Insomma ci sono storie di zombi perché ci sono anche storie dei fantasmi: le storie delle auto-zombi suonano invece strane perché non ci sono storie di auto-fantasma (ma ce n'è una sulle navi: l'*Olandese volante*). Se domani si scoprisse che gli stati mentali sono stati fisici del cervello allora fra qualche secolo i filosofi non potrebbero più riuscire a concepire menti separate dai corpi, così come, dopo le scoperte di Faraday, non è più concepibile che i metalli non conducano elettricità. La morale? Eccola: spesso l'immaginazione precede la conoscenza, ma a volte è bene che la segua. E con ciò concludo¹⁸.

NOTE

* In ciò che segue ho unito due tipologie di lezioni che ho tenuto nell'ambito di *Pianeta Galileo 2006*: una a Grosseto il 19 ottobre, dal titolo *Il problema mente-corpo*, l'altra a Cecina il 9 novembre e a Massa il 24 ottobre dal titolo *Matrix et alia: intelligenza artificiale e filosofia*. Propongo qui un unico testo poiché il contenuto di tutti gli incontri è stato sostanzialmente lo stesso; nelle manifestazioni di Cecina e di Massa al contenuto presentato a Grosseto sono stati aggiunti i riferimenti al cinema. Poiché nel più ci sta il meno la versione che segue è quella di Cecina e di Massa.

¹ Un'agile lettura sugli esperimenti mentali è [7]

² Sul rapporto fra cinema e filosofia cfr. [17], [2].

³ Una macchina del genere è anche lo spunto del film *La mosca* di D. Cronenberg (1986) tratto da un racconto di G. Langelaan.

⁴ Un'altra questione molto discussa nella filosofia ha a che fare con la seguente domanda: è possibile che 2+2 non faccia 4? Perché no? Questa domanda è dello stesso tipo di quest'altra: il rame potrebbe non condurre elettricità? E ancora: Se Napoleone non avesse perduto a Waterloo sarebbe sempre Napoleone? Un Napoleone vittorioso a Waterloo sarebbe un «Napoleone-vittorioso-a-Waterloo» o non sarebbe per niente Napoleone? Se non avessi partecipato al Pianeta Galileo – come relatore o come ascoltatore - sarei sempre io? Questioni del genere, che hanno a che fare con la cosiddetta logica modale, sono presenti in numerosi film, anche non appartenenti al genere fantascientifico, come *La vita è meravigliosa*, *Ritorno al futuro* e *Sliding Doors*. Ma su di essi non dirò nulla poiché sono oggetto dell'intervento di Sani (cfr. *infra* pag. 355).

⁵ Cfr. [14]

⁶ B. W. Aldiss, *I Supertoys che durano tutta l'estate*, *I Supertoys quando arriva l'inverno* e *I Supertoys nella nuova stagione*, pubblicati negli USA tra la fine degli anni Sessanta e i primi anni Settanta del secolo scorso. Il primo è stato tradotto in italiano nel 1999 sul «Corriere della Sera» del 27 novembre 1999. Adesso si possono leggere tutti nella trad. it. in [1].

⁷ Cfr. [4]. Il successo del film indusse Clarke a riscrivere il racconto in forma più ampia, cfr [5]

⁸ Ma la posizione di Turing era più articolata, perfettamente consapevole dell'importanza dell'obiezione secondo la quale «il solo modo per cui si potrebbe esser sicuri che una macchina pensa è quello di *essere* la macchina e di sentire se stessi pensare» ma anche del fatto che questa strada porterebbe al solipsismo: «secondo questa opinione la sola via per sapere che *un uomo* pensa è quella di essere quell'uomo in particolare. [...] Invece di discutere in continuazione su questo punto, è normale attenersi alla convenzione – suggerita dalla buona creanza – che ognuno pensi» [19] p. 169

⁹ Cfr. [16]. Per un panorama introduttivo alle varie teorie della mente cfr. [15]

¹⁰ *Viaggio allucinante* (1966) per la regia di R. Fleischer tratto da un racconto di O. Klement, si segnalò per gli effetti speciali e vinse due premi Oscar. I. Asimov trasse un racconto omonimo dalla sceneggiatura del film che uscì nello stesso anno e poi un *Viaggio allucinante II* intitolato *Destinazione cervello*.

¹¹ Ha scritto recentemente McGinn: «la proprietà della coscienza stessa (o di specifici stati consci) non è una proprietà osservabile o percepibile del cervello. Potete risiedere in un cervello conscio vivente, nel vostro o in quello di qualcun altro, e osservare una gran varietà di proprietà - forme, colori, strutture e così via - ma con ciò non vedrete quello che il soggetto sta sperimentando, cioè lo stesso stato conscio. Gli stati consci non sono semplicemente, in quanto stati consci, potenziali oggetti di percezione: essi dipendono dal cervello ma non possono essere osservati dal dirigere i sensi sul cervello. Non è possibile vedere uno stato cerebrale in quanto stato conscio». [12, p. 183. Traduzione modificata]

¹² Non ho prove che gli insetti non abbiano una coscienza ma credo che tutti lo pensino: se ci fosse qualcuno che pensa il contrario possibile che non abbia mai protestato per il genocidio che noi umani perpetrriamo tutte le estati a danno delle zanzare? L'attribuire una mente a un essere ha anche inevitabili implicazioni etiche e giuridiche. In uno dei suoi interventi filosofico-fantascientifici H. Putnam si chiede se i robot debbano godere dei diritti civili. Anche a questo riguardo si potrebbero citare numerosi film fra i quali lo stesso *AI* in cui David si trova al centro di un dilemma morale quando i genitori decidono di disfarsene. Un robot che rivendica e ottiene diritti civili è Andrew, protagonista di *L'uomo bicentenario* di C. Columbus (1999) tratto da un racconto di I. Asimov del 1976.

¹³ B. Libet ha condotto molti esperimenti che dimostrano come la presa di coscienza di un atto volontario segue di circa 250 ms i corrispondenti eventi cerebrali. Ciò significherebbe che noi non decidiamo nulla: è il nostro cervello che decide e noi (cioè la nostra coscienza) ne veniamo messi al corrente dopo. È una forma di epifenomenismo sostenuta da prove sperimentali. Tuttavia Libet riconosce alla coscienza un diritto di veto: essa può intervenire a bloccare una decisione già deliberata dal cervello. Si tratterebbe dunque di una forma di semi-epifenomenismo. Cfr. [11] e [8]. La metafora – ottocentesca - che paragona la coscienza al fumo di una locomotiva intende illustrare l'epifenomenismo: il fumo è causato dalla locomotiva ma non ha nessun effetto su di essa.

¹⁴ In realtà esistono molte versioni diverse del funzionalismo e dunque molte teorie computazionali. Teoria computazionale – al singolare – è dunque il nome di una famiglia.

¹⁵ Come si vede si tratta di una nozione perfettamente coerente con la descrizione che abbiamo dato del mal di denti.

¹⁶ L'astronomo F. Hoyle ha immaginato nel racconto di fantascienza *La nuvola nera* una nebulosa nella quale la materia è organizzata in modo da produrre intelligenza. Cfr. [10]. Non so se sia stato tratto un film da quel racconto, so che è stato tratto uno sceneggiato televisivo da un altro racconto di Hoyle intitolato *La nebulosa di Andromeda*. Ma molti anni fa quando la televisione era in bianco e nero e (ci crederete?) occorreva alzarsi dalla poltrona per cambiare canale

¹⁷ Naturalmente, come tutti gli altri, anche l'argomento della stanza cinese ha sollevato un vasto dibattito, ricordiamo qui soltanto la replica di Hofstadter il quale ha osservato che se è vero che la simulazione al calcolatore di un uragano non è un uragano, la simulazione di una dimostrazione matematica è una dimostrazione matematica [9, p. 99].

¹⁸ Una confessione finale: non è vero che ho conosciuto personalmente Kubrik.

BIBLIOGRAFIA

- [1] Aldiss B. W., *A.I. Intelligenza artificiale*, Mondadori, Milano 2001.
- [2] Cabrera J., *Da Aristotele a Spielberg. Capire la filosofia attraverso i film*, Bruno Mondadori, Milano 2003.
- [3] Chalmers D., *La mente cosciente*, McGraw-Hill, Milano 1999.
- [4] Clarke A. C., *La sentinella*, (1953), in «Gamma», n. 14, gennaio 1967.
- [5] Clarke A. C., *2001 Odissea nello spazio*, (1968), Longanesi, Milano 1972.
- [6] Descartes R., *Discorso sul metodo*, Laterza, Bari 1978.
- [7] Cohen M., *Lo scarabeo di Wittgenstein e altri classici esperimenti mentali*, Carocci, Roma 2006.
- [8] Gazzaniga M., *La mente inventata*, Guerini, Milano 1999.
- [9] Hofstadter D., Dennett D., *L'io della mente*, Adelphi, Milano 1985.
- [10] Hoyle F., *La nuvola nera*, Garzanti, Milano 1966.
- [11] Libet B., *Mind Time*, Cortina, Milano 2007.
- [12] McGinn C., *Possiamo risolvere il problema mente-corpo?* in [16] pp. 173-196.
- [13] Putnam H., *I robot: macchine o vita creata artificialmente?*, in *Mente, linguaggio e realtà*, Adelphi, Milano 1987, pp. 416-438.
- [14] Putnam H., *Cervelli in una vasca*, in *Ragione verità e storia*, Il Saggiatore, Milano 1985, pp. 7-28.
- [15] Salucci M., *Mente/Corpo*, La Nuova Italia, Firenze 1997.
- [16] Salucci M. (a cura di), *La teoria dell'identità. Alle origini della filosofia della mente*, Le Monnier, Firenze 2005.
- [17] Sani A., *Il cinema tra storia e filosofia*, Le Lettere, Firenze 2002.
- [18] Searle J., *Menti, cervelli e programmi. Un dibattito sull'intelligenza artificiale*, CLUP-CLUED, Padova 1984.
- [19] Turing A., *Macchine calcolatrici e intelligenza*, (1950), in Somenzi V. (a cura di), *La filosofia degli automi*, Torino, Boringhieri 1965, pp. 116-156.