



“Sei umana?” Hari Seldon e l’atteggiamento intenzionale

Persio Tincani*

1. Il dubbio di Seldon

Il *Ciclo della Fondazione* è una serie di romanzi di Isaac Asimov che narra una lunga storia di fantapolitica ambientata in un lontano futuro. Grazie alla tecnologia del viaggio iperspaziale l’umanità ha colonizzato l’intera galassia, organizzata come un enorme stato federale con capitale nel pianeta Trantor. L’origine comune dell’uomo da uno stesso pianeta è una tesi molto discussa, perché della Terra si è persa la memoria e nessuno, davvero, sa dire se si tratti di un pianeta esistente oppure di un racconto mitologico, un residuo delle religioni ormai scomparse da tempo. In un lontano passato gli uomini hanno inventato i robot, che via via sono stati perfezionati fino a divenire indistinguibili dagli umani. Questi, molto evoluti, sono stati dotati di cervelli positronici che garantiscono a essi una perfetta autonomia di comportamento: i robot sono in grado di prendere decisioni, di compiere scelte molto complesse, di dare consigli agli umani e così via. Essi, infatti, sono stati progettati e costruiti per aiutare l’umanità e per aiutare ogni singolo essere umano, e questa convivenza ha apportato grandi benefici. Il loro cervello funziona secondo le tre leggi della robotica, che stabiliscono i comportamenti che un robot deve sempre tenere e le cose che un robot non può fare, in nessun caso:

1. Un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno.
2. Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge.
3. Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Prima o con la Seconda Legge.

È importante chiarire subito che le Leggi non sono doveri in senso deontico, ma doveri anankastici. Nessun robot, infatti, è in grado di violare le Leggi, perché il loro cervello è progettato in modo da rispettarle sempre.

In un certo momento storico, i robot sono stati dichiarati nemici dell’umanità (non serve qui ricordare perché) ed è stato promulgato un decreto che ha impartito loro l’ordine di disattivarsi. La maggior parte dei robot ha obbedito a quest’ordine, eseguendo la seconda Legge, ma una parte dei robot non ha potuto farlo, perché si trovava nella condizione in cui eseguire la seconda Legge avrebbe violato la prima. Così,

* Persio Tincani è professore associato di Filosofia del diritto presso il Dipartimento di Giurisprudenza dell’Università di Bergamo.

nonostante il decreto (e soprattutto: senza violare le Leggi), qualche robot rimane in circolazione, in quanto impegnato in attività di protezione di esseri umani, o della specie umana. Sono stati i robot stessi, infatti, a integrare le Tre Leggi con una quarta legge, la Legge Zero, che costituisce il principio senza il quale le Leggi perderebbero gran parte del loro senso e della loro efficacia. In un dialogo con Hari Seldon, il robot Daneel (uno dei protagonisti della saga), così spiega questa integrazione delle Leggi:

Avevo un... un amico, ventimila anni fa. Un altro robot. Non come me. Era impossibile scambiarlo per un essere umano [...]. Secondo questo amico, doveva esserci una norma ancor più generale delle Tre Leggi... L'ha chiamata Legge Zero, dal momento che zero precede uno. È questa:

Legge Zero. Un robot non può fare del male all'umanità o, tramite l'inazione, permettere che l'umanità riceva danno.

Dunque la Prima Legge deve essere completata in questo modo: Un robot non può fare del male a un essere umano o, tramite l'inazione, permettere che un essere umano riceva danno, a meno che questo non contrasti con la Legge Zero. [Asimov 1991: 510; v. anche 1988: 294]

Il decreto rimane valido anche a distanza di secoli, così i robot devono condurre la propria esistenza senza rivelare di essere robot; cosa non semplice, dato che un corollario delle Leggi è che un robot non può mentire.

Il protagonista di gran parte del *Ciclo della Fondazione* è Hari Seldon, un professore universitario che ha inventato la psicostoria, una sorta di filosofia della storia meccanicistica che consente di conoscere gli eventi sociali in anticipo. La scienza di Seldon, è ovvio, può avere importantissimi effetti sulla vita politica e, pertanto, c'è chi vorrebbe ucciderlo per evitare che divulghi informazioni che sarebbero nocive per i propri piani. Per questa ragione, Seldon vive sotto protezione e, un bel giorno, gli viene assegnata come guardia del corpo la bellissima scienziata Dors Venabili, una donna dall'intelligenza eccezionale e dotata di forza e di capacità di combattimento fuori dal comune. Come è forse prevedibile, Seldon finisce per innamorarsi di lei e tenta qualche avance. C'è però un dubbio che lo tormenta: Dors è umana? Quella donna è troppo forte, troppo intelligente e, soprattutto, sembra non invecchiare mai. Nel finale di uno dei romanzi del *Ciclo*, assistiamo a questo interessante dialogo.

[Seldon] Esegui gli ordini e rischi la vita per me senza esitare, incurante delle conseguenze. Hai imparato a giocare a tennis in un attimo. Hai imparato a maneggiare i coltelli ancor più in fretta e nello scontro con Marron ti sei comportata in modo perfetto... in modo disumano, se mi consenti il termine. Hai una forza muscolare sorprendente, e reazioni di una rapidità sorprendente. Riesci a capire quando c'è qualcuno che origlia quel che si dice in una stanza, e riesci a metterti in contatto con Hummin [N.d.A: una delle false identità del robot Daneel] senza usare apparentemente alcuno strumento.

Dors disse: "E da tutto ciò che conclusioni trai?"

"Be', ho pensato che Hummin, nel suo ruolo di R. Daneel Olivaw, si trova di fronte a un'impresa impossibile. Come può un solo robot cercare di guidare l'Impero? Deve avere degli aiutanti."

"Questo è ovvio. Milioni di aiutanti, immagino. Io sono un'aiutante. Tu sei un aiutante. Il piccolo Raych, pure"

"Tu sei un aiutante di tipo diverso"

"In che senso? Hari, dillo. Se sarai tu a dirlo e lo sentirai con le tue stesse orecchie, ti renderai conto di quanto sia assurdo."

Seldon fissò Dors a lungo, poi fece sottovoce: “Non lo dirò... perché non mi importa”.

“Davvero? Vuoi prendermi come sono?”.

“Ti prenderò come devo. Tu sei Dors e, qualunque altra cosa tu sia, io non voglio nient'altro.”

Dors mormorò: “Hari, io voglio il tuo bene proprio perché sono quel che sono, ma sento che anche se fossi diversa vorrei ugualmente il tuo bene. E non penso di essere adatta a te”.

“Adatta o no, non m'importa.” Seldon abbassò lo sguardo e fece qualche passo, soppesando le sue prossime parole. “Dors, sei mai stata baciata?”

“Certo, Hari. Fa parte della vita sociale, e io ho una vita sociale.”

“No, no! Voglio dire, hai mai baciato veramente un uomo? Sai, appassionatamente?”

“Be', sì, Hari.”

“Ti è piaciuto?”

Dors esitò. “Quando ho baciato in quel modo, è stato più bello che deludere un uomo che mi piaceva, un uomo la cui amicizia significava qualcosa per me.” A questo punto, Dors arrossì e distolse lo sguardo. “Per favore, Hari, per me è difficile spiegare.”

Ma Seldon, più che mai deciso, insisté. “Dunque hai baciato per la ragione sbagliata... per evitare di ferire i sentimenti di qualcuno.”

“Forse lo fanno tutti, in un certo senso.”

Seldon rifletté su quelle parole, poi disse all'improvviso: “Tu hai mai chiesto di essere baciata?”

Dors rimase in silenzio alcuni istanti, come se stesse riesaminando la propria vita.

“No.”

“O dopo essere stata baciata, non hai mai desiderato che ti baciassero di nuovo?”

“No.”

“Hai mai dormito con un uomo?” chiese Seldon sottovoce, disperato.

“Certo. Te l'ho detto. Queste cose fanno parte della vita.”

Seldon strinse le spalle di Dors, come se volesse scuoterla. “Ma non hai mai provato il desiderio... il bisogno di quel tipo di intimità con una persona in particolare, una persona speciale? Dors, non hai mai sentito in te l'amore?”

Dors alzò la testa lentamente, in modo quasi mesto, e fissò Seldon negli occhi.

“No, Hari, mi dispiace.”

Seldon la lasciò andare, abbandonando le braccia lungo i fianchi, abbattuto.

Dors, delicatamente, gli posò la mano sul braccio e disse: “Lo vedi, Hari. In fondo non sono quello che vuoi”.

Seldon piegò il capo e fissò il pavimento. Esaminò il problema e cercò di pensare con razionalità. Poi rinunciò. Era qualcosa che voleva... la voleva, al di là di qualsiasi considerazione razionale. Alzò la testa. “Dors, cara... anche così, non m'importa.”

La cinse con le braccia e accostò la testa alla sua, lentamente, quasi si aspettasse che lei potesse ritirarsi, ma continuò ad attirarla a sé.

Dors non si mosse e lui la baciò... un bacio lungo e appassionato... E di colpo le braccia di Dors lo strinsero.

Quando infine Seldon si staccò, lei lo guardò col sorriso negli occhi e disse:

“Baciarmi ancora, Hari... Ti prego”. [Asimov 1991: 515-17]

A prima vista, il dubbio di Seldon consiste nel non sapere se Dors è umana oppure se si tratta di un artefatto: se Dors è un artefatto, allora non è viva e, di conseguenza, non può essere umana. La questione, in realtà, è un po' più complessa, perché trascura il fatto ovvio che non tutte le cose vive sono umane. Nel caso di Dors, però, sembra che ciò non sia un punto rilevante: Dors si comporta come un essere

umano, perciò se è viva è sicuramente umana. Direi, dunque, di cominciare a tentare di risolvere la questione se Dors sia viva o no, e di riservare per un secondo momento l’analisi della questione se sia anche umana.

2. Che cosa è vivo?

Fino a poco tempo fa, rispondere alla domanda “Che cos’è la vita?” non costituiva un problema. La vita, si diceva, è “materia animata”, dal latino anima. Questa risposta, com’è evidente, non era affatto una spiegazione, poiché si limitava ad attribuire all’anima, ovvero allo spirito vitale, tutte le caratteristiche della vita che non erano state comprese. [de Duve 2003: 1]

Il punto è che, di primo acchito, siamo più o meno tutti convinti che distinguere le cose vive dalle cose inanimate sia piuttosto facile, anzi che sia talmente facile da non meritare neppure troppe riflessioni. Una macchina, come un robot, è un artefatto, e gli artefatti non sono vivi. Il computer sul quale sto scrivendo non è vivo, così come non lo è la rivista che avete in mano in questo momento; non è viva la sedia sulla quale sedete, né lo è la lampada sulla vostra scrivania (che, a sua volta, non è viva). Gli esseri umani, i cani, i gatti, i vermi, gli uccelli e così via sono vivi, mentre non lo sono le automobili e le biciclette, le navi, i carrarmati, i satelliti artificiali e così via. Tuttavia, sebbene i contenuti di queste affermazioni siano considerati dalla maggior parte di noi come intuitivamente veri – e veri in maniera indiscutibile – le cose si fanno più complesse se ci mettiamo a riflettere su che cosa ci rende così tanto certi che lo siano. Infatti, non esiste una definizione soddisfacente di “vita”, cioè una definizione che non sia fondata su affermazioni aprioristiche, almeno in qualche misura. Si potrebbe obiettare che la vita è qualcosa di organico, mentre le cose non vive sono inorganiche. Ma sarebbe sbagliato, ovviamente, perché se così non fosse dovremmo concludere che, per esempio, i tavoli di legno sono vivi, oltre a sollevare molti dubbi sul fatto che siano vive le stesse cellule (come quelle che compongono il nostro corpo), perché in ultima analisi tutti gli elementi che le compongono sono riducibili a complesse combinazioni di elementi inorganici.

Gli scienziati ancora oggi usano il termine chimica organica, anche se sappiamo che le leggi della chimica sono le stesse sia per una molecola che si trova dentro a un organismo sia per una che ne sia al di fuori. [Davies 2012: 32]

Del resto, è inorganico lo stesso carbonio, senza il quale la vita (almeno su questo pianeta) sarebbe impossibile. Anzi, non ha neppure senso dire che il carbonio (o qualsiasi altro elemento) è inorganico, perché gli elementi non sono né organici né inorganici: sono elementi e basta (e il carbonio è semplicemente carbonio).

Sarebbe inutile cercare una risposta guardando alla biologia, la scienza della vita per eccellenza, perché nemmeno la biologia ha mai formulato una definizione di “vita”, né è attrezzata per farlo. La biologia studia gli esseri viventi, tuttavia non li definisce: il biologo sa perfettamente (almeno, cerca di sapere perfettamente) quali sono i funzionamenti di un determinato essere vivente o di un’intera specie, ma questo non significa affatto definire la vita e, già che ci siamo, nemmeno definire che cosa sia una *specie* (e, infatti, non possediamo una definizione nemmeno di questo).

Si potrebbe rispondere che, sebbene non sia possibile fornire una definizione, è lo stesso possibile tracciare una linea di demarcazione tra ciò che è vivo e ciò che non lo è. Una linea di demarcazione promettente potrebbe essere l’affermazione che “in qualsiasi

luogo, la vita deve consistere di entità dal metabolismo autonomo” (Dennett 1997: 160). Tuttavia, neppure questa strategia è soddisfacente, in quanto consiste anch’essa in una presa di posizione convenzionale. Infatti, come nota ancora Daniel Dennett, coloro che definiscono la vita in questo modo

possono escludere i virus dalle forme viventi, pur tenendo i batteri nella magica cerchia. Può darsi che esistano buone ragioni per un decreto definitorio come questo, ma ritengo che si possa cogliere meglio l’importanza del metabolismo autonomo se lo si considera come una condizione profonda, se non del tutto necessaria, per quella sorta di complessità che si richiede per schivare gli effetti corrosivi del secondo principio della termodinamica. Tutte le strutture macromolecolari tendono a rovinarsi nel corso del tempo, quindi, a meno che un sistema non sia *aperto*, capace di assumere dentro di sé altre sostanze e riempirsene, tenderà ad avere una carriera molto breve. [Dennett 1997: 160-1]

In altre parole, possedere un metabolismo autonomo è, casomai, una proprietà che rende la vita (molto) più lunga, ma non è una proprietà indispensabile alla vita. Per rendere l’idea, in linea di massima è piuttosto vero che gli esseri dotati di capacità di movimento abbiano una vita più lunga degli esseri che ne sono privi, ma nessuno direbbe che la capacità di movimento è necessaria alla vita, dato che conosciamo molti esseri viventi che ne sono privi. Si potrebbe obiettare che, però, nessun essere vivente è privo del metabolismo autonomo, ma non si tratterebbe di una vera obiezione, in quanto si limiterebbe a ripetere l’assunto convenzionale “vita=metabolismo autonomo” senza fornire nessun nuovo argomento a suo sostegno.

Un candidato piuttosto promettente per aiutare nel definire un confine tra vita e non-vita potrebbe essere individuato nel DNA, ma in realtà nemmeno il DNA ci può servire a un gran che, visto che non tutti i virus sono provvisti di DNA (e, in effetti, sebbene sia chiaro che il virus è un’entità biologica, è discusso se si tratti di un’entità viva oppure no). E anche se dovessimo accettare che il DNA è un componente necessario di un organismo *vivo*, ci resterebbe il compito tutt’altro che semplice di spiegare il perché lo sia. Infatti, sostenere semplicemente che “la presenza di DNA in un’entità la qualifica come viva” senza specificare la ragione per la quale lo sosteniamo, avrebbe lo stesso valore del sostenere che “la presenza di uranio in un’entità la qualifica come viva”, oppure – perché no – che “la presenza di carta stagnola (o di una biglia di vetro rosso) in un’entità la qualifica come viva”. Ovviamente, nessuno crede che il possesso di DNA sia la condizione necessaria (o, addirittura, la condizione necessaria e sufficiente) per la vita, o che il DNA abbia proprietà magiche, né che si tratti di qualcosa di simile a un bollino di qualità. Piuttosto, la tesi si incentra sul fatto che il DNA è, in sostanza, una catena di informazione che ha la caratteristica di contenere le informazioni necessarie a essere replicata (de Duve 2003: 27 e ss).

Tuttavia, anche questa strada non porta molto lontano: la capacità di replicare sequenze e informazioni non è una caratteristica dei soli sistemi che definiamo viventi, e di conseguenza dovremmo accettare di definire vivi oggetti che nessuno considera tali, come i cristalli, per esempio. La sola “informazione”, dunque, non basta:

è ovvio che un ruolo fondamentale, in questa storia, lo gioca la complessità. Un sistema, per qualificarsi ad essere descritto come *vivente*, deve fare qualcosa in più rispetto a limitarsi a replicare l’informazione (questo lo può fare un semplice cristallo di sale): deve essere abbastanza complesso da possedere un po’ di autonomia. In altre parole, il contenuto di informazione deve essere abbastanza grande da far sì che questo sistema possa occuparsi dei propri programmi, ossia

che possa, abbastanza alla lettera, “prendere vita”. È ben lungi dall’essere chiaro cosa potrebbe essere questa soglia della complessità, anche se i microbi autonomi più semplici che si conoscano in natura contengono, ciascuno, oltre un milione di bit di informazione. [Davies 2012: 33]

Si potrebbe, così, ritenere che la vita stia nella quantità e nella qualità di informazione. Gli atomi contengono le informazioni necessarie per unirsi con altri atomi e formare molecole, e contengono anche le informazioni necessarie per ripetere questa operazione fino a quando non incontrano un segnale di stop. Attenzione, sia chiara una cosa: gli atomi non sanno niente! Ma, del resto, non sa niente nemmeno il DNA, eppure, come scrive con grande efficacia Richard Dawkins, tutti danziamo alla sua musica (Dawkins 1995: 131).

Tuttavia, nemmeno la quantità (e la qualità) di informazione può davvero dirci qualcosa di soddisfacente in merito alla questione vita / non-vita. Certo, *per il momento* funziona bene per quanto riguarda i cosiddetti “organismi superiori”, come i vertebrati e come gli stessi esseri umani, perché in questi casi la quantità e la qualità di informazione sono davvero impressionanti. Per dare un’idea,

ciascuna cellula del nostro corpo contiene l’equivalente di quarantasei lunghissimi nastri di dati – i cromosomi – che via via rilasciano caratteri digitali attraverso numerose testine di lettura simultaneamente operanti. I cromosomi di ciascuna cellula contengono le stesse informazioni, ma le testine di lettura dei diversi tipi di cellule cercano parti diverse del *data base* per i propri scopi particolari. Questo è il motivo per cui le cellule dei muscoli sono diverse dalle cellule del fegato. Non vi è nessuna forza vitale metafisica, e non vi è alcuna mistica gelatina protoplasmatica che vibri, si agiti o ribolla. La vita è semplicemente byte e byte e byte di informazioni digitali. [Dawkins 1995: 28]

Il genoma umano è complessivamente composto di circa 3.200.000.000 (tre miliardi e duecento milioni) di coppie di basi di DNA, ciascuna delle quali contiene circa 25.000 geni. Con gli altri animali superiori, le cose non vanno in modo molto diverso. Si tratta di una quantità di informazione impressionante; e si tratta di informazione di *ottima* qualità, perché complessivamente gli organismi sono ben funzionanti e le specie riescono a perpetuarsi nel tempo. *Per il momento*, ripeto, nessun artefatto contiene abbastanza informazione da poter essere anche lontanamente paragonata a questa enorme quantità.

Tuttavia, non per tutti gli esseri viventi le cose vanno in questo modo. Per esempio, nel caso dei batteri, il DNA è portatore di una quantità di informazioni *molto* più modesta, nell’ordine di 4.67 Mb, eppure nessuno discute sul fatto che i batteri siano vivi. Bene, allora dobbiamo chiarirci per quale ragione non riteniamo che i nostri computer anch’essi siano vivi, visto che contengono molta più informazione di quella che è contenuta in qualsiasi batterio, che nella migliore delle ipotesi arriva a occupare lo spazio che occupa qualche file di testo salvato sul nostro hard disk (il pdf dell’ultimo mio libro, che non arriva a trecento pagine, è di circa 1.4 Mb). L’informazione genetica contenuta nella drosofila (il moscerino della frutta) è di circa 180 Mb, più o meno quanto le dimensioni di un semplice programma per computer.

Se accettiamo che la strategia della quantità di informazione sia quella giusta per tracciare la linea vita/non-vita, allora dobbiamo per forza di cose accettare anche una di queste conclusioni: a) il nostro computer è vivo; b) il nostro computer non è vivo e non lo sono nemmeno i batteri e nemmeno un sacco di altre cose che, intuitivamente,

consideriamo vive, ma che hanno un contenuto di informazione più ridotto di quello dei nostri computer.

Apro qui un inciso, perché è necessaria una precisazione. Quando mi riferisco all'*informazione*, intendo l'informazione espressa dal genoma e l'informazione espressa dalla stringa di scrittura del programma. Questo è il solo modo in cui le due categorie (programmi e organismi) possono essere messe a confronto: il numero di bit del quale consistono i rispettivi codici. Oltre questo livello, le cose vanno in maniera diversa nei due casi: nel caso del programma, la singola unità di informazione è determinata dalla scelta tra due possibilità (gli 0 e 1 del sistema binario), mentre nel caso del genoma la singola unità di informazione è determinata dalla scelta e dalla combinazione tra 4 possibilità (Adenina, Timina, Citosina e Guanina – ATCG). Le possibilità combinatorie, quindi, sono molto più elevate nel secondo caso rispetto al primo. Per la precisione, quelle del secondo caso sono il quadrato di quelle del primo: un programma di 16 Mb è l'esito di un processo di possibilità di informazione di 16^2 Mb; il genoma del batterio di 4.67 Mb è, invece, l'esito di un processo di possibilità di informazione di 4.67^4 Mb.

Può essere che queste conclusioni suonino stonate perché, in fondo, siamo convinti che i computer non siano vivi e, invece, siamo convinti che lo siano i batteri. Ma per supportare questa nostra convinzione intuitiva, allora, siamo costretti ad abbandonare la strategia della quantità di informazione, perché è la sua peggior nemica. Confrontare i computer con i batteri e con altri organismi semplici può creare qualche imbarazzo, ma possiamo ancora provare a nascondere sotto il tappeto. Tuttavia, le cose sono destinate a farsi sempre più imbarazzanti: guardando a quanto è accaduto negli ultimi decenni, è ragionevole prevedere che lo sviluppo delle tecnologie produrrà computer sempre più potenti in tempi molto brevi, così che la quantità di informazione contenuta in ciascuno di essi sarà ancora più grande e di qualità sempre migliore (quando scrivevo la tesi di laurea, nei primi anni Novanta del secolo scorso, disponevo di un computer con un hard disk di 20 Mb e con 2 Mb di RAM, e si trattava di una macchina dell'ultima generazione. La capacità di memoria del mio attuale telefonino è di 32 Gb, e in giro c'è di meglio). Se vogliamo continuare a considerare i computer non vivi e, nello stesso tempo, tenerci saldi alla strategia della quantità di informazione, allora dobbiamo prepararci a provare degli imbarazzi ben più gravi di quelli che ci causano adesso i batteri, perché la tecnologia informatica e la cibernetica sono destinate a spostare l'assicella sempre più in alto.

3. In che senso esiste la vita?

Spesso le domande filosofiche sono quelle che mettono in discussione l'ovvio, insinuando il dubbio fastidioso che ovvio non sia affatto. La domanda che rileva qui, e che è sottesa a tutte le considerazioni che ho fatto in precedenza, non è, come potrebbe forse sembrare, "Come si fa a distinguere tra vita e non-vita?", ma "*Esiste* una distinzione tra vita e non-vita?"

Ovviamente, questa distinzione esiste, perché la effettuiamo di continuo. Si tratta, per di più, di una distinzione che funziona piuttosto bene nel quotidiano, perché nella maggior parte dei casi ci troviamo tutti d'accordo su che cosa deve essere definito 'vivo', perciò si tratta di una distinzione molto solida e molto ben funzionante nel contesto della normale comunicazione. Tuttavia, non appena proviamo a capire su che cosa essa si fondi, cioè su quali siano le basi oggettive sulle quali è costruita la linea di demarcazione (che pure utilizziamo) vita/non-vita, ci rendiamo conto che queste basi

non sono altrettanto solide. Vita e non-vita sono qualità che attribuiamo agli oggetti esterni (e a noi stessi) senza una ragione oggettiva. Oggettivamente, la sola cosa che possiamo vedere è che ci sono cose che posseggono un codice genetico e cose che non lo posseggono; che ci sono codici genetici che incorporano una quantità di informazione più elevata di quella contenuta in oggetti che non posseggono un codice genetico, ma che è vero anche il contrario; che ci sono cose nelle quali gli elementi costitutivi sono combinati in un certo modo e cose nelle quali gli stessi elementi sono combinati in un altro modo, e così via. Tutti questi sono dati oggettivi, che non è possibile discutere, ma che non ci dicono assolutamente niente in merito a che cosa sia o non sia vivo. Accanto a questi dati, e indipendentemente da essi, c’è un altro dato oggettivo: di alcune cose diciamo che sono vive, e ci troviamo tutti d’accordo; di altre diciamo che non lo sono e ci troviamo, ancora, tutti d’accordo.

Il fatto è che la vita non è qualcosa come una mela, come un computer o come un temporale, qualcosa della quale posso dire con certezza se sia o meno davanti ai miei occhi in questo momento. Quando guardo le mie gatte che si rincorrono in giardino, non vedo “la vita” ma vedo due gatte, un prato, degli alberi e dei sassi (e un rastrello che, con tutta evidenza, è stato abbandonato lì da mia figlia quando si è stancata di giocarci). Si dirà che è evidente che le gatte, gli alberi e il prato sono qualcosa di vivo, e che è altrettanto evidente che i sassi non lo sono. Bene: è vero, è evidente e io non lo metto in dubbio. Tuttavia, è evidente “soltanto” sul piano delle credenze ma, come abbiamo visto, cessa di essere evidente quando ci si colloca su di un livello di osservazione più profondo, perché quel livello, semplicemente, ignora la distinzione tra vita e non-vita.

La domanda filosofica: “*Esiste una distinzione tra vita e non-vita?*”, dunque, non ha una risposta secca. La risposta è: “No, non esiste” se ci collochiamo sul livello di osservazione che potremmo chiamare della “causalità profonda”; ma diventa subito: “Certo che esiste!”, se ci collochiamo sul livello di osservazione delle credenze o, per essere più precisi, della causalità che inferiamo dai dati dei quali abbiamo esperienza quotidiana e diretta¹. Si dirà che una di queste risposte è per forza sbagliata, in quanto una contraddice l’altra. In realtà, le cose non stanno proprio così, perché le due risposte si collocano su due piani diversi, che non sono davvero in contatto e che, pertanto, non hanno molte possibilità di entrare in contraddizione. Distinguere tra vita e non-vita è impossibile sul livello dell’osservazione profonda, perché quel livello è puramente descrittivo, mentre quella distinzione è tutta costruita sul piano valutativo. Cercare di spingere la distinzione fino al livello profondo sarebbe, perciò, impossibile, oltre che di nessuna utilità. Distinguere tra vita e non-vita, diventa invece possibile ed estremamente utile sul piano del quotidiano (per esempio, pensate all’utilità di sapere che non possiamo essere aggrediti da un’automobile in sosta e che, invece, possiamo essere aggrediti da un cane, anche se nel momento in cui li vediamo l’una e l’altro sono perfettamente fermi) e infatti è una distinzione che effettuiamo di continuo e che, sul livello di osservazione normale, funziona benissimo. Il livello profondo – quello che ho

¹ L’idea dei diversi livelli di causalità risale a David Hume ([1748] 1996: 97 ss.), che riteneva che fosse impossibile conoscere l’autentico nesso causale che collega due eventi l’uno all’altro, in quanto l’attribuzione di un nesso causale è sempre compiuta sulla base del livello di conoscenza della realtà che abbiamo raggiunto, senza che sia però possibile sapere se il succedersi degli eventi non sia in realtà causato da fatti dei quali non sappiamo niente a causa di una nostra imperfetta conoscenza della realtà. Sui diversi piani di conoscenza causale e sul pensiero di Hume in merito, mi permetto di rimandare a Tincani 2004: 35-8.

chiamato altrove *causalità-1* (Tincani 2004) – non mi dice niente sui cani aggressivi, perché i cani aggressivi esistono su un altro livello, quello della *causalità-2*, che esperisco nel quotidiano. Il cane è vivo (e l'automobile non lo è) soltanto su questo secondo livello, che è un livello che ho interamente strutturato sulla base di credenze che ho costruito in maniera valutativa. Si tratta di altrettante letture simboliche della realtà nella quale siamo calati e della quale noi stessi siamo parte e, ciò che più conta, si tratta di letture che non trovano alcun riscontro e nessuna fondatezza non appena abbandoniamo il livello della *causalità-2* per un livello che entra in maggiore profondità nella materialità degli eventi.

Causalità-1 e *causalità-2* sono entrambe descrizioni del medesimo oggetto che, di quello stesso oggetto, ci dicono cose diverse e in apparente contraddizione; “ciò non crea però nessun problema, dal momento che la *causalità-2* offre una chiave di lettura del mondo che funziona benissimo, anche se non è necessariamente vera” (Ivi: 37); inoltre, come abbiamo visto, la contraddizione è soltanto apparente, perché non vi può essere contraddizione tra un'affermazione valutativa e un piano di analisi (la *causalità-1*) interamente descrittivo nel quale non esiste spazio per la valutazione.

Può darsi che, in questo caso, l'espressione 'letture simboliche' appaia opaca. Tuttavia, non c'è un altro modo per descrivere il passaggio dell'attribuzione della qualità della vita a oggetti del mondo (compresi noi stessi) che sono interamente costituiti da elementi e da istruzioni rispetto ai quali non ha neppure senso porsi la questione vita/non-vita. La distinzione esiste sul livello di *causalità-2* in quanto questo livello è quello che si è formato per effetto dell'evoluzione dei nostri cervelli, un'evoluzione che procede mediante la conservazione di mutazioni che si sono rivelate positivamente adattive. Il buon funzionamento di questa distinzione è spiegabile nei termini del *prezzo metabolico*: ogni capacità cognitiva, come ogni capacità in generale, ha un prezzo da pagare. Lo scoiattolo che sta rosicchiando una ghianda fugge appena sente un rumore e abbandona lì il suo pasto. Noi che osserviamo la scena pensiamo che lo scoiattolo fugga perché teme che quel rumore potrebbe essere il passo di un predatore e vuol mettersi in salvo. Certo, quel rumore potrebbe anche essere il suono di un grosso frutto che cade a terra – un pasto ben migliore della ghianda – ma lo scoiattolo preferisce applicare il *maximin* e, fatto un rapido calcolo dei costi e dei benefici della coppia di possibilità predatore/frutto, scappa: perdere di sicuro la ghianda (e forse la ghianda e il grosso frutto) non è un risultato ottimale, ma è di sicuro migliore di quello di perdere la vita.

È ovvio che gli scoiattoli non sappiano niente di *maximin* (ma non è grave, perché nemmeno la maggior parte degli esseri umani ne ha la minima idea) e che, perciò, non è questa la ragione per la quale si mette in fuga lasciando il pranzo a metà. La ragione è che il comportamento “fuggi quando senti un rumore” si è rivelato più vantaggioso, dal punto di vista adattivo, del comportamento “vai a vedere se c'è qualcosa da mangiare quando senti un rumore”. In un contesto nel quale i predatori esistono, gli animali portatori del meme “rumore–*slurp*” hanno una ragionevole prospettiva di carriera nell'ambiente solo se sono, appunto, i predatori. Negli altri casi, quel comportamento si potrebbe rivelare *molto* controproducente (e nel lungo periodo, prima o poi, lo farà), così che quell'animale finirà per essere mangiato. Per dirlo con delle metafore, la selezione naturale ha *scartato* (ecco una metafora) gli animali che non si sono dimostrati in grado di reagire con efficienza agli stimoli dell'ambiente, *creando* (un'altra metafora) dei comportamenti automatici che migliorano le loro possibilità di sopravvivenza con il minimo costo. Nel caso del cervello, questo costo è lo zucchero. Compiere una rappresentazione mentale è laborioso e costoso, ma per fortuna la selezione naturale ci ha messo nelle condizioni di non aver sempre bisogno di costruirne una nuova. Così,

i nostri antenati non avevano bisogno di sapere che la rappresentazione di un orso era in quel momento attiva nel loro cervello o che in quel momento stavano prestando attenzione a uno stato interno rappresentante il lento avvicinarsi di un lupo. Nessuna delle due immagini richiedeva così loro di bruciare zucchero prezioso. Tutto quello che dovevano sapere era: “Orso là!” oppure “Un lupo si sta avvicinando da sinistra”. Non era necessario per la sopravvivenza riconoscere che questo era soltanto un *modello* del mondo e dell’“ora”. Questo tipo di conoscenza addizionale avrebbe richiesto la formazione di quelle che i filosofi chiamano *metarappresentazioni*, o immagini che raffigurano altre immagini, pensieri che vertono su altri pensieri. Ciò avrebbe richiesto hardware addizionale nel cervello e più carburante. L’evoluzione talvolta produce casualmente nuovi tratti superflui, ma raramente queste lussuose proprietà sono conservate per un lungo periodo di tempo. [Metzinger 2010: 51]

Allo stesso modo, sul livello della *causalità-2* la costruzione simbolica della “vita” funziona perfettamente, perché ci consente di orientarci in modo molto efficiente in un ambiente nel quale può essere vantaggioso comportarci con ciò che classifichiamo come vita in maniera diversa da come ci comportiamo con ciò che classifichiamo come non-vita.

4. Umani e macchine

Sono seduto al tavolino di un bar quando scorgo, dall’altra parte della strada, la statua di un faraone dell’antico Egitto con le mani incrociate sul petto. Le persone che passano lì davanti si fermano e la osservano con divertito interesse, il che è per me incomprensibile dato che in quella statua non c’è davvero niente di così notevole da giustificare il capannello. Poi, guardando meglio, vedo che la statua si muove: in modo molto lento, con piccoli scatti, il braccio destro si sposta dal petto e si protrae di fronte, avvicinando agli spettatori quello strano scettro somigliante a un bastoncino dolce di Natale che stringe in pugno, per poi compiere il movimento al contrario fino a tornare nella posizione iniziale. Qualcuno fa cadere qualche moneta in una scodella posta ai piedi della statua e si allontana divertito. La statua continua con quello stesso movimento per un po’, attirando altri spettatori che si comportano più o meno come quelli di prima: osservano divertiti e qualcuno di loro lascia una moneta nella scodella. Quella statua è un automa, questo l’ho capito, però non mi pare che sia un automa in grado di far niente di eccezionale. Muovere un braccio lentamente non richiede tutta questa capacità di progettazione. Mia figlia ha molte bambole che fanno di meglio: per esempio, ha un Little Pony di nome Rainbow Dash che fa dei perfetti salti mortali all’indietro, un Ciccibello che fa le bolle di sapone, un altro Ciccibello che cammina (e che sa anche gattonare), dei criceti coloratissimi che trainano una carrozza e che, se li passi sopra a una chiave magica, si mettono a ballare, e ha un sacco di altri giocattoli che fanno fare molte più cose di quella statua di dubbio gusto (io stesso possiedo un robot *molto* più sofisticato di quel faraone: un’aspirapolvere che va da sola! La programmo e quella, all’orario stabilito, parte e pulisce il pavimento, evitando gli ostacoli e tornando da sé alla base quando ha finito oppure quando ha bisogno di ricaricarsi). Dopo un po’, però, succede qualcosa che mi spiega il perché di tutto quell’interesse da parte degli spettatori. La statua distende le braccia, scende dal piedistallo e cammina fino a una panchina poco distante, si siede, si toglie la maschera, tira fuori una sigaretta e se l’accende. Non è un automa progettato in modo da fingere di essere un uomo, ma un uomo che finge di essere un automa! Devo ammettere che è anche piuttosto bravo: lo

spettacolo è di una noia mortale e non spenderei una lira per assistervi, ma credo che stare immobili per un'ora o più, muovendo soltanto il braccio destro in quel modo sempre uguale, non dev'essere per niente facile.

Come ho potuto pensare che si trattasse di un automa? Beh, perché *faceva cose da automa*. Il suo movimento era ripetitivo, sempre lo stesso e, soprattutto, non sembrava rivelare nessuna intenzione da parte della macchina. Non si notava nessun indizio di interazione con gli spettatori, niente, per intenderci, di quello che fanno di solito gli artisti di strada, che ammiccano, sorridono al pubblico e spesso coinvolgono qualcuno per giochi di prestigio o per qualche sketch. In poche parole, faceva una cosa troppo semplice per lasciar supporre che si trattasse di un essere umano. Una volta scoperto il gioco, però, la situazione si ribalta; lo spettacolo è tutto basato proprio su questo: “sono un essere umano” (a quanto pare, ero io il solo a non saperlo) “ma mi comporto in modo tale da creare una sorta di spaesamento, perché mi comporto come una macchina semplice. Sono tanto più bravo – e tu sei tanto più divertito – quanto più sono semplici le cose che faccio, quanto più non-umani sono i miei movimenti” (in effetti, nei primi tempi in cui questa forma d'arte di strada si cominciava a vedere in giro, gli attori stavano perfettamente immobili, cosa che gli è valso il nome di “statue viventi” con le quali sono adesso chiamate).

Proviamo adesso a pensare a un caso molto diverso. Sono di nuovo seduto al tavolino di un bar (va bene, è un periodo in cui ho molto tempo libero) e si avvicina un signore che regge un uccellino colorato sull'indice. L'uccellino muove gli occhi in quella maniera caratteristica che hanno gli uccellini e cinguetta come un usignolo. In un tavolo vicino al mio c'è un bambino seduto con i genitori, e il signore con l'uccellino si dirige proprio lì. Il bambino osserva divertito: quell'uccellino gli piace davvero tanto e i genitori glielo comprano. Il signore prende il denaro e consegna l'uccellino al bambino che lo stringe tra le mani tutto contento. Il venditore estrae da una borsa un altro uccellino, se lo appoggia sul dito e questo, solo a quel punto, comincia a cinguettare. Per contro, l'uccellino del bambino rimane perfettamente zitto, ma questo non sembra rappresentare per il piccolo nessun problema. Il bambino, al contrario di me, aveva capito subito tutto: quello non è un vero uccellino, ma un pupazzo, un giocattolo con gli occhi semoventi (gli occhi sono due semisfere di plastica trasparente con dentro una pallina nera che si sposta quando il giocattolo viene mosso). In effetti, non è l'uccellino a cinguettare, ma il venditore, che conosce qualche trucco da ventriloquo e lo usa per pubblicizzare la merce in un modo simpatico che, a quanto pare, sembra funzionare. Perché il venditore è riuscito a ingannarmi per un po', facendomi credere che quello fosse *davvero* un uccellino? Beh, perché *faceva delle cose da uccellino*. Non è affatto straordinario vedere, per esempio, dei passerotti che stanno fermi sul posto e cinguettano muovendo gli occhi. Il gioco è finalizzato a farci pensare questo:

- Che bello, quel signore tiene in mano un uccellino che cinguetta.
- Ora che guardo meglio, non è un uccellino, ma un pupazzo. È *lui* che cinguetta senza muovere le labbra.
- Uh, quanto è bravo questo signore simpatico.
- Beh, carino però questo uccellino. Lo compro.

Anche in questo caso si verifica un effetto di spaesamento, ma per ragioni opposte a quello che si registra nel caso della statua vivente perché opposto è l'inganno che sta alla base dello show. Nel primo caso, un oggetto vivente viene dissimulato in un oggetto non-vivente; nel secondo, è un oggetto non-vivente viene dissimulato in un oggetto vivente. In entrambi i casi, l'inganno si realizza mediante l'esecuzione di

comportamenti tipici: è tipico dell’automa, non dell’essere umano, muoversi in quel modo; è tipico degli uccellini veri, non dei pupazzi, muoversi in quel modo.

L’inganno che viene messo in atto dalla statua vivente e dal ventriloquo consiste nel fingere di essere qualcosa d’altro. Noi, che siamo le vittime dell’inganno, ci caschiamo o meno a seconda di quanto sia verosimile la rappresentazione. Ma attenzione: non c’è nulla di genuinamente fraudolento nei due casi. Il ventriloquo non dice, per esempio, “questo è un passerotto vero”, ma si limita a fischiettare tenendo in mano un pupazzo; altrettanto, la statua vivente si limita a muovere un braccio in modo che anche noi, con un po’ di allenamento, saremmo in grado di fare: ciò che fa, in questo senso, è del tutto umano (del resto, la statua vivente è un essere umano). Siamo noi, in altre parole, che collaboriamo con loro per creare l’inganno, perché cataloghiamo ciò che vediamo davanti a noi come ‘macchina’ o come ‘uccellino’ secondo una classificazione che abbiamo costruito sulla base dell’osservazione di comportamenti che abbiamo già visto in passato (“veri” passerotti e “veri” automi). In definitiva: nessuno racconta frottole, ma alla fine noi siamo lo stesso ingannati.

Il celebre Test di Turing funziona sullo stesso principio. Siamo in grado di distinguere una macchina da una persona sulla base delle risposte che l’una e l’altra forniscono alle nostre domande? (Turing 1950, v. anche 1994). La tesi sostenuta da Turing è, in breve, che se non siamo in grado di farlo significa che quella macchina è in grado di pensare. Il Test di Turing ha innescato un ampio dibattito molto fertile che continua tuttora (Gillies 1998). Alcuni, come John Searle (1980), hanno sostenuto che nessun test sia in grado di fornire la risposta a una domanda del genere, altri hanno proposto delle modifiche del test per renderlo meno esigente (o, viceversa, più esigente). Possiamo però lasciare tutto questo da parte e concentrarci su una domanda che il test non affronta: come posso sapere se è in grado di pensare l’essere umano coinvolto nel test di Turing, quella persona che devo provare a distinguere dalla macchina sulla base delle risposte che dà alle mie domande? In quella situazione, dove le sole informazioni delle quali dispongo sono le risposte che mi vengono date, la realtà è che non posso davvero saperlo. La sola cosa che posso fare è comparare le risposte che provengono dal soggetto A (l’umano) con quelle che mi vengono date dal soggetto B (la macchina) e dire se le prime e le seconde sono diverse sotto aspetti rilevanti per poter affermare che il soggetto A è diverso dal soggetto B. Si pensi a un caso pratico di test di Turing, il gioco on line, per esempio il gioco degli scacchi. Un nostro amico si siede al computer ed entra in un sito dove è possibile giocare a scacchi con un avversario che si trova da un’altra parte del mondo. Sul monitor appare una scacchiera, dove i pezzi sono mossi selezionandoli con il mouse e spostandoli. Come può fare a sapere se l’avversario è umano e non è, invece, un programma che gioca a scacchi? Il giocatore umano potrebbe ricorrere ad alcuni indizi. Per esempio, sta perdendo una partita dopo l’altra da mezza giornata; questo potrebbe far supporre che l’avversario, che non sembra diminuire mai la concentrazione e che non sbaglia nessuna mossa, non sia una persona in carne e ossa, perché prima o poi le persone si distraggono e prima o poi una mossa la sbagliano. In realtà, però, questo indizio non è così indicativo come potrebbe sembrare. Prima di tutto, l’avversario potrebbe essere semplicemente un giocatore di scacchi più bravo del nostro amico. Anche la considerazione che “non sbaglia nessuna mossa” mentre prima o poi le persone lo fanno non ci consente di andare più in là, perché è ben possibile che l’avversario dall’altra parte del mondo, in effetti, *sbagli* alcune mosse (anche un numero alto di mosse) ma che il nostro amico non sia abbastanza bravo da accorgersene. A pensarci bene, però, anche se questi indizi non vi fossero non sarebbe possibile capire se dall’altra parte ci sia un giocatore umano o un programma. Mettiamo che il nostro amico vinca qualche partita perché riesce ad accorgersi di qualche mossa sbagliata del

suo avversario. Questo significa che ha trovato un avversario alla sua portata, ma ciò può avvenire tanto nel caso che dall'altra parte ci sia un giocatore umano, quanto nel caso che dall'altra parte ci sia un programma settato su un livello di abilità medio-basso (anche i programmi di scacchi più semplici, che possiamo scaricare gratuitamente da internet, offrono la possibilità di selezionare diversi livelli di difficoltà). Un altro indizio potrebbe essere dato dalla velocità delle mosse dell'avversario, ma anche questo non aiuta un gran che. Prima di tutto, l'avversario potrebbe essere più bravo (ancora) del nostro amico, conoscere molte partite e, pertanto, muovere più velocemente senza doverci pensare troppo su. In secondo luogo, nel caso che l'avversario muova velocemente ma poi finisca per perdere la partita, questo potrebbe significare sia che si tratta di un giocatore umano distratto e poco abile, sia che si tratti di un programma settato su un basso livello di difficoltà.

Potremmo andare avanti a ipotizzare altri possibili indizi che possano suggerire al nostro amico una risposta alla domanda "Sto giocando contro un essere umano o contro un programma?", ma non sarebbe molto utile perché credo che sia ormai chiaro che non esiste un metodo universale per capirlo (né un indizio-principe). Al massimo, possono esistere situazioni concrete nelle quali gli indizi sono piuttosto forti, ma anche in questi casi non si può davvero dire che essi ci consentano di andare più il là del sospetto (magari anche di un forte sospetto). Un giocatore piuttosto abile potrebbe essere sorpreso dal fatto di non essere mai riuscito a vincere una partita contro l'avversario on-line, e perciò potrebbe sospettare di avere a che fare con un ottimo programma settato su un livello di difficoltà molto elevato; ma è ovvio che, per quanto sia improbabile, non può escludere di trovarsi alle prese con Magnus Carlsen.

Finora abbiamo considerato soltanto alcune delle possibili situazioni nelle quali il nostro amico scacchista potrebbe trovarsi, ma non è affatto detto che le cose debbano andare così. Potrebbe capitare, infatti, che il nostro amico vinca una partita dopo l'altra. Il suo avversario sbaglia tutte le mosse, commette gli errori del principiante, compresa la suicida apertura Barnes che conduce in modo pressoché inesorabile al *matto dell'imbecille*. In questo caso, sospetterebbe di aver a che fare con un programma? Immagino di no. Ma la domanda, a questo punto, è: perché se il tuo avversario gioca da cani è normale che tu non sia sfiorato dal dubbio di star giocando contro un programma?

Per certi versi, quando ci sediamo alla scacchiera e cominciamo a giocare a scacchi non ci comportiamo in modo diverso da come ci comportiamo in ogni altro contesto caratterizzato dall'interazione. In ogni caso, ci formiamo delle aspettative sul comportamento degli altri, perché nelle interazioni il nostro atteggiamento è caratterizzato dal fatto che compiamo delle previsioni su ciò che gli altri faranno. Ciò non significa che siamo sempre certi di che cosa gli altri faranno, cioè che siamo sempre in grado di prevedere con esattezza il loro comportamento, ma che compiamo previsioni che sono in grado di restringere in maniera drastica l'incertezza sulle possibili azioni future altrui, considerando soltanto le più probabili. Questo atteggiamento è noto come "atteggiamento intenzionale", e consiste nell'operazione di attribuire intenzionalità agli oggetti esterni con i quali entriamo in relazione o dei quali assumiamo contezza. La sua scoperta si deve a Daniel Dennett, che riguardo al caso di questo esempio scrive:

ho sempre sottolineato l'effettivo potere predittivo del puro atteggiamento intenzionale. Ho sostenuto, per esempio, che è possibile utilizzare l'atteggiamento intenzionale per prevedere il comportamento di un ignoto avversario (umano o artificiale) nel gioco degli scacchi, e questo suscita talvolta l'obiezione che ciò che fa degli scacchi un gioco interessante è proprio il fatto che le mosse del proprio avversario *non* sono prevedibili adottando l'atteggiamento intenzionale.

Questa obiezione fissa standard di previsione troppo elevati. Innanzitutto, esistono quelle situazioni a fine partita in cui parliamo di “mosse obbligate”, che sono prevedibili adottando l’atteggiamento intenzionale (e soltanto l’atteggiamento intenzionale) con un margine effettivo di sicurezza del 100 per 100 [...]. Le mosse obbligate sono obbligate soltanto nel senso di essere “dettate dalla ragione” [...].

Che dire delle mosse dell’avversario a metà partita? Queste di rado sono prevedibili in modo attendibile e univoco [...], ma è una situazione rara quella in cui le 30 o 40 mosse *legali* di cui può disporre l’avversario non possano essere ridotte, tramite l’atteggiamento intenzionale, a un breve elenco disordinato di una mezza dozzina di mosse più probabili, su cui si potrebbe scommettere con successo, dando alla pari tutte le mosse legali. Ciò comporta un enorme vantaggio predittivo che si fonda su esili elementi, a dispetto della pressoché totale ignoranza dei meccanismi che entrano in gioco, grazie al potere dell’atteggiamento intenzionale. [Dennett 1993: 115-6]

In sostanza, l’atteggiamento intenzionale è un modo quasi completamente pragmatico di relazionarsi con la realtà circostante mediante l’attribuzione di una qualità agli oggetti. Una qualità che attribuiamo anche a noi stessi – l’intenzionalità – e che consiste nella capacità di possedere stati mentali relativi a qualcosa (non necessariamente esistente) nel mondo. Questi stati mentali possono essere legati causalmente a qualcosa, ma non è necessario. L’attribuzione di intenzionalità presuppone, a sua volta, l’attribuzione di (almeno una forma di) autocoscienza: io posso pensare che è da troppo tempo che non suono il mio contrabbasso perché ho sì la capacità di pensare al trascorrere del tempo, all’attività di suonare e a un contrabbasso (oltre che a una serie di altre cose come, per esempio, gli effetti che derivano dal trascurare la pratica di uno strumento musicale), ma anche perché, prima di tutto, ho la capacità di pensare a *me* come a un soggetto distinto da tutto quello che ho adesso menzionato e, in ultimo, di avere coscienza di tutto ciò: il contrabbasso, il tempo, l’attività di suonare e io che penso a queste cose. “Aristotele e Franz Brentano hanno sottolineato in maniera simile che il percepire in maniera cosciente implica la consapevolezza del fatto di percepire qualcosa in maniera cosciente, ora, in questo dato istante. In un certo senso, dobbiamo percepire l’atto di percepire mentre sta accadendo” (Metzinger 2010: 35; v. anche Brentano [1867] 1989). L’atteggiamento intenzionale consiste nell’attribuire questa stessa capacità, anche se non necessariamente lo stesso grado di abilità nel riconoscimento di oggetti o di eventi, alle cose che esistono nel mondo esterno. L’intenzionalità, a sua volta, è il presupposto per avere intenzioni. Perciò, l’atteggiamento intenzionale si può ridurre, con una semplificazione che non causa nessun problema sul livello empirico delle normali situazioni di interazione o della relazione con oggetti esterni, all’attribuzione di *intenzioni* – ovvero *desideri* – agli oggetti del mondo esterno. Nel caso del nostro amico e della sua ormai interminabile serie di partite a scacchi, l’atteggiamento intenzionale consiste nell’attribuire all’avversario la catena di intenzioni: a) intenzione di giocare a scacchi, b) intenzione di vincere la partita, c) intenzione di compiere ogni singola mossa in modo funzionale alla soddisfazione dell’intenzione/desiderio b). In poche parole, l’atteggiamento intenzionale funziona così: “innanzitutto si stabilisce di trattare l’oggetto di cui si deve prevedere il comportamento come un agente razionale; poi si individuano le credenze che l’agente in questione dovrebbe avere, data la sua posizione e il suo scopo nel mondo. Quindi si determinano i desideri che dovrebbe avere, sulla base delle stesse considerazioni, e infine si può prevedere che questo agente razionale agirà per favorire i suoi obiettivi alla luce delle sue credenze”. (Dennett 1993: 33)

Come mai, allora, se il nostro amico infila una vittoria dopo l’altra perché il suo avversario gioca in modo pessimo, siamo portati a pensare che abbia a che fare con un

avversario umano e non con un programma di scacchi? In fondo, il programma potrebbe essere costruito male, e quindi formulare le intenzioni c) in modo non funzionale alla realizzazione dell'intenzione b); oppure, il programma potrebbe essere costruito in modo di non avere l'intenzione b) “vincere la partita” ma l'intenzione b) “perdere la partita”.

La risposta è semplice: perché i programmi di scacchi non hanno queste caratteristiche. Se abbiamo giocato contro un programma (una volta o spesso, non importa), sappiamo che vincere non è facile perché il programma è costruito per giocare bene (più o meno bene, a seconda del livello di difficoltà impostato e, ovviamente, a seconda di quanto è sofisticato); magari siamo giocatori molto bravi, e troviamo che i programmi giochino male (non ci piace il loro stile, per esempio), ma nemmeno in questo caso pensiamo che i programmi giochino così tanto male da commettere gli errori del principiante con disarmante frequenza. Nessuno giocherebbe con un programma così, perché accenderlo non vorrebbe dire “Adesso gioco una partita” ma nient'altro che (e sempre) “Adesso vinco”. Questo vale anche sia per chi non ha mai giocato in vita sua a scacchi contro un programma, sia per chi non ha mai giocato a scacchi *tout court*; e vale anche per chi ha appreso in questo momento che esiste un gioco che si chiama ‘scacchi’, del quale ignora tutto (dalle regole alla modalità materiale del suo svolgimento), e che esistono programmi contro i quali è possibile giocare a scacchi. L'atteggiamento intenzionale è un sistema per attribuire credenze (in questo caso: “credo che questa mossa sia la migliore che potrei fare”) a un sistema che ci raffiguriamo come dotato di intenzionalità (la capacità di avere credenze), ma questa attribuzione non viene fatta alla cieca. Al contrario,

la prima regola per attribuire credenze nella strategia intenzionale è la seguente: attribuire come credenze tutte quelle verità rilevanti per gli interessi (o desideri) del sistema che l'esperienza del sistema medesimo ha finora reso disponibili. [Dennett 1993: 34]

Ciò non esclude che un sistema possa avere false credenze, ma se abbiamo sufficiente esperienza del sistema questo non è molto rilevante per la nostra capacità di prevederne i comportamenti. E la nostra esperienza dei sistemi giocatore umano / programma di scacchi ci dice che la normalità è che le false credenze del tipo “farmi mangiare il più presto possibile le torri e gli alfieri è un'ottima scelta per vincere la partita” non sono le credenze che potremmo attribuire a un programma di scacchi e, pertanto, non è quello che ci aspettiamo che il programma farà. Una credenza del genere (è un esempio estremo di proposito: in realtà, non credo che nemmeno un principiante possa avere esattamente questa credenza) è invece quella che potremmo supporre in un principiante assoluto, perciò, se il nostro amico si troverà davvero in quella situazione (vincere tutte le partite contro un avversario che sembra non avere la minima idea di come si giochi), sarà l'atteggiamento intenzionale che gli farà concludere che il suo avversario è umano.

Le credenze c) sembrano a posto, nel senso che l'atteggiamento intenzionale le spiega bene in termini di corrette informazioni e di corretta razionalità strumentale per il conseguimento di uno scopo desiderato che attribuiamo al sistema. Come possiamo, però, sapere che il sistema ha proprio *quello* scopo e non un altro? Come facciamo, insomma, a sapere che il nostro avversario – umano o programma che sia – ha lo scopo di vincere la partita e non, invece, quello di perderla? Anche in questo caso, l'atteggiamento intenzionale funziona sulla base di un'analogia con noi stessi:

come attribuiamo i desideri (preferenze, obiettivi, interessi) sulla cui base formuleremo poi l’elenco delle credenze? Noi attribuiamo i desideri che il sistema *dovrebbe avere*. È questa la regola fondamentale. Essa impone, come primo punto, di attribuire alle persone la serie, a noi familiare, dei desideri più importanti o fondamentali: sopravvivenza, assenza di dolore, cibo, comfort, procreazione, divertimento. Citare uno di questi desideri significa, in particolare, chiudere il gioco che consiste nel dare una ragione alla domanda “perché?”. Non dovrebbe esserci più bisogno di un’ulteriore motivazione per desiderare il comfort o il piacere o il prolungamento della propria esistenza. Le regole derivate dall’attribuzione di desiderio interagiscono con le attribuzioni di credenza. In parole povere, abbiamo la seguente regola: bisogna attribuire desideri di quelle cose che un sistema crede siano positive per se stesso. [Dennett 1993: 36-7]

Se mi siedo per giocare a scacchi, il primo desiderio che ho è quello di giocare, il che implica che io voglia vincere (se mi siedo con l’intenzione di perdere, non sto giocando a scacchi, ma sto facendo un altro gioco: per esempio, voglio trarre piacere dal fatto che l’avversario si senta appagato). L’atteggiamento intenzionale attribuisce lo stesso desiderio anche all’avversario, quindi l’intenzione b) “vincere la partita” è la stessa per me e per lui.

Alla fine, non potremo mai essere certi di aver riconosciuto la macchina e l’umano, perché anche l’atteggiamento intenzionale, almeno in questo caso, è collocato sul livello della *causalità-2*, caratterizzato da ridotte informazioni e dal ricorso a sistemi simbolici che consentono classificazioni utili all’economia delle nostre decisioni. Possiamo riconoscere che è molto improbabile che un programma di scacchi sia costruito in modo da giocare malissimo, ma non possiamo escludere che sia così. Allo stesso modo, se siamo campioni di scacchi e non siamo mai riusciti a battere quell’avversario, riconosceremo che è molto più probabile che dall’altra parte ci sia un ottimo programma rispetto all’eventualità che ci sia Anatolij Evgen’evič Karpov, oppure Garry Kasparov. Tuttavia, non è possibile escludere nemmeno questa possibilità. Magari dall’altra parte c’è davvero Kasparov. E magari c’è sempre lui anche nel primo caso, quello dell’avversario scadente, e si sta divertendo a giocare male per ragioni sue, non importa sapere quali. Il livello di *causalità-2* non è perfetto, come non lo è l’atteggiamento intenzionale. Ciò che conta, però, è che dimostrino di funzionare bene per i nostri scopi, ed è fuor di dubbio che lo facciano *molto* bene. Essi sono stratagemmi selezionati dall’evoluzione in modo da consentire il c.d. “pensiero veloce” (Dennett 1993: 385 e ss.) (è una metafora, ovviamente: l’evoluzione non seleziona proprio niente, e tantomeno lo fa per qualche scopo), un vantaggio evolutivo enorme in termini di costo metabolico. Come per il concetto di vita / non-vita, tutto dipende dal livello sul quale ci poniamo e dagli obiettivi che vogliamo realizzare. L’atteggiamento intenzionale non è perfetto, ma ci consente di indovinare – in particolari condizioni e con particolari conoscenze (come la nostra abilità di scacchisti, la probabilità che abbiamo di incontrare un giocatore con le caratteristiche del nostro avversario, la nostra normale conoscenza del fatto che gli esseri umani tendono prima o poi a distrarsi o a commettere un errore, e così via) – se *in casi specifici* abbiamo a che fare con giocatore umano o con una macchina, con un margine di errore tollerabile che tendiamo a non considerare. Tuttavia, l’atteggiamento intenzionale non può dirci in che modo sapere sempre se abbiamo a che fare con giocatori umani o con macchine, per le stesse ragioni per le quali non può farlo nessuna strategia costruita sulla *causalità-2*: su questo livello di causalità, la sola cosa che le tecniche di ragionamento possono fare è fornire una risposta che consideriamo soddisfacente per i nostri scopi su questo livello causale, e non esiste (senza cambiare livello causale) nessuna possibilità di andare oltre a questo.

Bene, a questo punto il nostro amico ha capito: dall'altra parte c'è una macchina. Si è fatto tardi, così spegne il computer e se ne va a letto. L'indomani mattina, mentre beve il caffè, lancia ancora il programma per il gioco on-line e trova una situazione ribaltata: vince con una facilità disarmante due partite e, quando vede che anche la terza sta prendendo la stessa piega dopo poche mosse, abbandona il gioco maledicendo i ragazzini che, senza la minima conoscenza del gioco, si mettono a cercare avversari su internet con il solo risultato di far perdere agli altri del tempo prezioso. Spegne tutto e corre in stazione: ha fatto tardi e rischia di perdere il treno che lo porterà al lavoro. Non saprà mai che ieri sera ha giocato con Karpov e che stamattina ha giocato con un programma costruito per fare imbestialire questi spocchiosi scacchisti.

5. *You don't know what love is*

Il caso che ho discusso nel paragrafo precedente è un esempio di come potrebbe essere svolto un Test di Turing in condizioni nelle quali esistono delle regole che stabiliscono di che tipo devono essere le domande da porre al soggetto del quale si deve cercare di capire se è umano o se è una macchina. Il gioco originario, quello dal quale Turing ha sviluppato il suo test, consiste nel cercare di indovinare se una persona è un uomo o una donna sulla base delle risposte che dà alle domande del giocatore (per iscritto o mediante un intermediario). La prima regola riguarda il personaggio misterioso, e stabilisce che le risposte devono essere sincere. Le altre sono rivolte al giocatore, che non può domandare qualunque cosa. Ovviamente, non può domandare "Sei un uomo?", perché il gioco finirebbe subito (anzi, non sarebbe nemmeno un gioco); inoltre, non può domandare niente la cui risposta svelerebbe in modo inequivocabile il sesso del personaggio misterioso ("Hai il pene o la vagina?", per esempio, può essere considerata equivalente alla prima domanda vietata). Il gioco può prevedere la presenza di un arbitro che decida di volta in volta se la domanda è ammessa o no, perché in effetti è molto difficile tracciare a priori delle regole che stabiliscano quali domande non possono essere poste in quanto la risposta equivarrebbe a una confessione. Per esempio, la domanda "Hai allattato al seno i tuoi figli?", che pure a prima vista può sembrare appartenente al gruppo delle domande-killer, non è in realtà sempre così definitiva: la risposta "Sì" identificherebbe di certo il personaggio misterioso come una donna, ma la risposta "No" sarebbe quella fornita da una donna senza figli, da una donna che non ha allattato i propri figli al seno e, naturalmente, da qualsiasi uomo.

Si può notare con facilità che tanto più sono severe le regole che determinano quali domande sono ammesse, quanto più è difficile indovinare se il personaggio misterioso è un uomo o una donna (se potessi domandare tutto, farei subito la domanda "Sei un uomo?"). Se non è possibile fare domande che si riferiscono ad attività esclusivamente maschili o femminili, le cose si fanno ancora più complesse. Per esempio, le regole potrebbero escludere le domande che fanno riferimento ad attività esclusivamente maschili o esclusivamente femminili, come l'allattamento; se il personaggio misterioso è una donna che ha allattato al seno, la risposta equivale a una confessione. O, un tempo, come il servizio militare: la leva obbligatoria riguardava soltanto i maschi, non tutti i maschi avevano svolto il servizio militare, ma se il personaggio misterioso è un uomo che ha svolto il servizio militare non potrà che rispondere con l'equivalente di una confessione alla domanda "Hai fatto il militare?" o "Dove hai fatto il militare?" (la risposta "A Cuneo" equivale alla dichiarazione "Sono un uomo"). Nel caso dello scacchista, le interazioni ammesse con l'altro giocatore (le

domande e le risposte) sono determinate in modo tassativo e sono molto poche, consistendo soltanto nelle mosse consentite dalle regole degli scacchi, alle quali si possono aggiungere (ma non è qualcosa che aiuti molto) le considerazioni che il giocatore compie fondandosi sulla propria esperienza di gioco in passato (la velocità della mossa dell'avversario, la sua capacità di mantenere alto il livello di attenzione, la frequenza delle vittorie e delle sconfitte e così via). In queste condizioni, il gioco (quello di capire se l'avversario è umano o no) può davvero ridursi a tirare a indovinare, anche perché le mosse dell'avversario non devono per forza essere sempre l'equivalente delle risposte sincere delle regole del gioco originario (una mossa sbagliata può essere considerata, in certe condizioni, l'equivalente di una menzogna: dovrei muovere il pezzo X e vincerei in sei mosse, ma non posso farlo perché sono settato su un livello di difficoltà intermedia – oppure, nel caso di un giocatore umano: voglio farti giocare ancora un po' per farti allenare – e muovo il pezzo Y).

Il caso di Seldon e di Venabili è ancora diverso. Si tratta, infatti, di un specie di Test di Turing con regole, dove però le regole prevedono soltanto limitazioni tassative delle risposte e lasciano le domande del tutto libere. Seldon può domandare quello che vuole, mentre un robot non può rispondere qualunque cosa. Prima di tutto, un robot non può mentire (come non può mentire il personaggio misterioso del gioco originale) e deve eseguire gli ordini di un essere umano (quindi, non può nemmeno evitare di rispondere). Questo potrebbe sembrare una porta aperta verso la possibilità che la domanda-killer ("Sei un robot?") debba ricevere per forza una confessione, se l'interlocutore lo è, dato che un robot non potrebbe che rispondere la verità, e che dunque a Seldon basterebbe porla per togliersi ogni dubbio, ma le cose non stanno così. Prima di tutto, perché la regola di dire la verità non vale in assoluto ma vale soltanto per i robot, che sono costruiti in modo da non poter mentire. Se Venabili fosse un essere umano potrebbe mentire a piacimento, rispondendo "Sì, sono un robot" per osservare le reazioni di Seldon, oppure per sfidarlo, oppure per punirlo per quella domanda insultante ("Cosa diavolo è venuto in mente a questo imbecille! Un robot? Io?").

In realtà, dal dialogo tra Dors e Seldon traspare con evidenza che lui, Seldon, è praticamente certo che Dors sia un robot. La sua forza fisica, la sua rapidità di apprendimento, la capacità di comunicare con un altro robot senza usare nessuno strumento: tutto questo non lascia molte incertezze sul fatto che quella donna sia un artefatto. Quello che Seldon vuol sapere – e questo è il suo vero dubbio – è se quell'artefatto sia davvero indistinguibile da un essere umano dal punto di vista di tutti gli aspetti rilevanti per lui: la sincerità dei sentimenti, la spontaneità dei suoi gesti d'affetto e, in generale, se lei sia affezionata a lui in modo sincero e spontaneo e non a causa delle Tre Leggi. Così, rivolge a Dors una serie di domande per cercare di capire quale sia la causa dei comportamenti affettuosi – umani – di Dors. È indubbio che quella macchina imiti perfettamente un essere umano e che passerebbe qualsiasi Test di Turing. Ma questo a Seldon non basta, vuol sapere perché; o meglio: *desidera* che Dors gli dia una risposta che soddisfi il suo desiderio di amare essendo *davvero* corrisposto. Così, la interroga incalzante per sentirsi dire che sì, la ragione per le azioni di Dors è qualcosa di analogo, di indistinguibile da un sentimento.

Vediamo questo altro dialogo, questa volta tra il robot Daneel e Gladia (un'abitante del pianeta Solaria):

"Sei contento di rivedere Elijah Baley?"

"Non saprei come descrivere con precisione il mio stato interiore, signora. Forse è analogo a quello che gli esseri umani definiscono contentezza."

"Però proverai pure qualcosa, no?"

“Ho la sensazione di riuscire a prendere le mie decisioni più rapidamente del solito; sembra che le mie reazioni giungano con maggior facilità, che i miei movimenti richiedano meno energia. Potrei interpretarla complessivamente come una sensazione di benessere. Almeno, ho sentito usare questa parola dagli esseri umani e credo indichi qualcosa di simile a quanto sto provando.” [Asimov 1988: 34]

Proviamo a pensare a questo: rivolgeremmo mai domande del genere a un essere umano? Probabilmente no, perché non avrebbe nessun senso fargliele. Quello che tanto Seldon quando Gladia stanno facendo è cercare di ottenere una definizione razionale dei sentimenti, per vedere se alla base di quelli che i robot mostrano di provare con i loro comportamenti vi sia un sentimento “vero” o una stringa di istruzioni che modificano lo stato interno della macchina per farle compiere determinate azioni e non altre.

Il punto è che questo tentativo di indagine razionale dei sentimenti è destinato a dare delle grosse delusioni a chi voglia trovare un discrimine tra il sentimento “vero” umano e quello “artificiale” robotico, e ciò per le stesse ragioni che abbiamo visto quando abbiamo cercato di tracciare un confine tra vita e non-vita. Proviamo a pensare alla domanda: perché ci innamoriamo? Una prima risposta è quella più familiare, che fa riferimento a comportamenti o a caratteristiche delle persone (gli occhi della persona amata, la sua sensualità, la sua gentilezza e così via), e che spesso ci piace definire, almeno in parte, come qualcosa di irrazionale (“In fondo nessuno sa perché ci si innamora, e questa è una delle cose belle dell’amore!”). Una seconda risposta è decisamente meno romantica, e fa riferimento alle cause materiali della sensazione di innamoramento, in ultimo riducibili a una complessa serie di istruzioni che passano, sotto forma di stimoli elettrochimici, lungo la rete neurale. È importante notare che le due risposte non si contraddicono l’una con l’altra: esse si riferiscono a due diversi livelli di osservazione (due diversi piani causali) e sono entrambe vere. La seconda è del tutto razionale, la prima no. La seconda spiegazione, quella delle istruzioni lungo la rete neurale, ci dice qualcosa che può sembrare sgradevole, se non stiamo attenti a non mescolare i piani dell’osservazione: non siamo davvero liberi di innamorarci, perché la sensazione soggettiva di innamoramento (che comprende le famose farfalle nella pancia) è tutta provocata da una catena causale di eventi che, una volta innescata, è puramente deterministica. Le cose vanno in modo diverso sul livello della prima spiegazione (che chiameremo il *livello dell’amore romantico*):

quando si dice che l’amore è cieco, lo si dice senza dolersene. Si ritiene che l’amore debba essere cieco: l’idea stessa di una valutazione è bandita, quando si tratta di vero amore. Ma perché? Il senso comune non dà risposta, e gli economisti più testardi hanno da tempo bollato la cosa come insensatezza romantica, ma l’economista evoluzionista Robert Frank ha sottolineato che, in realtà, esiste un eccellente fondamento razionale (libero) per il fenomeno dell’amore romantico nel mercato sregolato degli umani giochi di corteggiamento. “Poiché la ricerca ha costi elevati, è razionale fissarsi su un partner prima di aver esaminato tutti i candidati potenziali. Ma spesso, una volta che un partner è stato scelto, le circostanze relative cambiano [...]. L’incertezza che ne deriva rende imprudente fare investimenti comuni, che sarebbero nel puro interesse di ciascuno dei due partner [...] per facilitare tali investimenti, ciascuna parte vuole sancire un impegno vincolante a rimanere nella relazione [...]. Le caratteristiche personali oggettive possono continuare a svolgere un ruolo nel determinare quali persone siano inizialmente attratte l’uno dall’altra, come l’esperienza insegna. Ma i poeti hanno senz’altro ragione quando dicono che il legame che chiamiamo amore non nasce da decisioni razionali basate sulla considerazione di quelle caratteristiche. Si tratta, piuttosto, di un legame intrinseco, in cui la persona è valutata per se stessa. Ed è proprio qui che risiede il suo valore

di soluzione del problema dell'impegno." [Dennett 2007: 273, cita Frank 1988: 195-6]

Questo significa che non esiste quello che chiamiamo "innamoramento", oppure che i nostri sentimenti sono sempre insinceri e frutto di un calcolo? Niente affatto; significa, invece, che l'innamoramento e il sentimento dell'amore sono un regalo dell'evoluzione; beninteso: un regalo interessato, perché l'evoluzione non dà mai qualcosa in cambio di niente. Come nota con grande chiarezza Dan Dennett, quando osserviamo i comportamenti o quando studiamo un organo o una funzione, una buona strategia è sempre quella di domandarsi "Cui prodest?", chi o che cosa trae vantaggio dall'esistenza di quella funzione, di quell'organo, di quel comportamento? A volte il vantaggio è per l'organismo che tiene quel comportamento (o che possiede quell'organo o che è capace di quella funzione), a volte il beneficiario del vantaggio va cercato da qualche altra parte: nella specie alla quale quell'organismo appartiene o addirittura in una specie diversa (ammesso che si sappia che cosa sia, una specie) (Dennett 2007: 3-7).

Nel nostro caso, ancora una volta, il vantaggio è il risparmio nel bilancio metabolico. Sebbene tutti conoscano casi nei quali l'innamoramento non ha portato ad altro che a una perdita di tempo e a un certo numero di sbronze per stemperare il dolore dei reiterati rifiuti, su larga scala (che è la sola scala che viene tenuta in conto dall'evoluzione) innamorarsi è un'ottima idea evolutiva: su larga scala, il meme dell'innamoramento apporta un *grande* vantaggio, perché abbatte i costi della ricerca del partner e crea un legame con esso che consente di compiere "investimenti comuni" come riprodursi, per esempio.

Può essere che questa spiegazione possa sembrare stonata, ma credo che ciò dipenda in gran parte dal fatto che l'amore è un sentimento che gode di ottima stampa in tutti i codici morali, e che pertanto si possa essere tentati a considerarlo qualcosa che, su ogni livello di spiegazione, sfugge alla razionalizzazione. Inoltre, le pene d'amore sono troppo pesanti, troppo reali per pensare che non siano tenute in nessun conto nel calcolo del vantaggio evolutivo che assegna all'amore il ruolo di fortunata scorciatoia verso il guadagno di un vantaggio adattivo. Proviamo allora a pensare a un altro sentimento, che per ragioni che mi sono oscure non gode di altrettanto favore: l'attrazione sessuale, riprovata dalla maggior parte dei codici morali religiosi. Non ci voleva David Hume per dirci che fare sesso è un'attività estremamente appagante, e sono sicuro che le persone lo trovassero emozionante anche prima che Hume scrivesse che è l'attrazione sessuale quello che fa sì che le persone sviluppino l'amore l'una verso l'altra. La molla che dà origine alle famiglie, il movente originario, è il sesso: alle persone piace fare sesso e dall'attività sessuale nascono i figli, verso i quali si sviluppa un sentimento di protezione (Hume [1739] 2001: 961).

Hume non sapeva niente di teoria dell'evoluzione, che ai suoi tempi ancora non esisteva, e forse è per questo che non nota che la sua argomentazione presenta un punto debole: se fosse semplicemente così, infatti, nessun sentimento d'amore potrebbe nascere tra persone che non sono in grado di fare sesso, per esempio a causa di impedimenti fisici legati all'età, dato che il suo sorgere è subordinato al sorgere della pulsione sessuale e dal desiderio di riuscire a soddisfarla. È la teoria dell'evoluzione che ci fornisce la spiegazione – e che ci dice che la spiegazione di Hume, benché incompleta, è tutto sommato corretta – perché ci impone di ragionare in termini di vantaggi evolutivi su una scala molto più ampia, che va al di là del singolo vantaggio che deriva a singoli individui. L'innamoramento è uno stratagemma evolutivo per indirizzare la pulsione sessuale, che a sua volta è uno stratagemma evolutivo per spingere alla riproduzione. È bello innamorarsi ed è bello fare sesso, ma non c'è nulla di intrinsecamente bello né in

una cosa né nell'altra. Possiamo immaginare con facilità un diverso esito evolutivo nel quale riprodursi non dia alcun piacere in nessuna fase del processo, perché è proprio quello che succede nel caso di molti animali non umani. Ma le trame dell'evoluzione, per quello che ci riguarda, hanno preso un'altra strada e hanno favorito gli individui che, provando piacere nel fare sesso, hanno finito per riprodursi di più. Il vantaggio evolutivo non è la possibilità di riprodursi, ma il fatto che riprodursi implichi delle attività piacevoli, anche *molto* piacevoli. L'amore è l'altro grande vantaggio evolutivo, in questo caso, perché riesce a interrompere la ricerca del partner inviando una sorta di segnale di stop. Sapere perché tutto questo succede, però, non interessa al nostro amore e non interessa alla nostra voglia di fare sesso. Allo stesso modo, il vantaggio adattivo che comportano l'una e l'altra sensazione non viene modificato dai singoli casi nei quali innamorarsi è una pessima idea e non porta a niente se non a sofferenza individuale:

Un uomo e una donna intelligenti leggono Darwin e sanno che il fine ultimo dei loro desideri sessuali è la procreazione; sanno che la donna non può concepire perché prende la pillola anticoncezionale, tuttavia questa consapevolezza non diminuisce in alcun modo il loro desiderio. Il desiderio sessuale è il desiderio sessuale, e la sua intensità, nella psicologia di un individuo, è indipendente dalla finalità darwiniana che lo suscita. È una pulsione forte che esiste a prescindere dalla sua fondamentale finalità biologica.

Sto suggerendo che lo stesso vale per la pulsione della bontà, ovvero dell'altruismo, della generosità, dell'empatia, della pietà. In epoca ancestrale, avevamo modo di essere altruisti solo verso i parenti stretti e i potenziali restitutori di favori. Oggi queste limitazioni non esistono più, ma la regola empirica continua a esistere. Perché non dovrebbe? È come il desiderio sessuale. Non possiamo fare a meno di provare pietà quando vediamo un infelice che piange (benché non sia nostro parente né possa restituirci favori), così come non possiamo fare a meno di provare desiderio per una persona [...] (che può essere sterile o comunque inidonea alla procreazione). Entrambi i sentimenti sono lacune, errori darwiniani; errori benedetti e preziosi.

[...] *Per la selezione naturale, il modo migliore di introdurre entrambe le pulsioni in epoca ancestrale era inscrivere regole empiriche nel cervello. Queste regole ci influenzano tuttora, anche nelle circostanze che le rendono inappropriate alle funzioni originarie.* [Dawkins 2007: 219-20, corsivo mio]

Dire che l'amore è un trucco dell'evoluzione non è molto romantico, bisogna ammetterlo, ma nessuno pretende che lo sia perché l'amore romantico non è da cercare qui, ma altrove. Come scrive Steven Pinker (1997: 418),

mormorare al vostro innamorato che il suo aspetto, la sua capacità di reddito e il suo quoziente d'intelligenza incontrano i vostri standard minimi probabilmente ammazzerebbe ogni alone di romanticismo, benché l'affermazione sia statisticamente vera. La via per giungere al cuore di una persona è dichiarare l'opposto: che siete innamorati perché nessuno può farci niente. [anche in Dennett 2007: 273-4]

Come per il concetto di vita/non-vita, anche l'amore romantico non esiste in ogni livello di spiegazione causale, ma soltanto sul livello della mente simbolica (quel "soltanto", sia chiaro, non deve in nessun caso essere inteso nel senso di voler sminuire l'importanza di nessuno dei due elementi, vita/non-vita e amore romantico). Anzi, ho forti dubbi che si possa plausibilmente parlare di "amore" su un livello di spiegazione causale diverso da quello del piano della mente simbolica. È soltanto nel regno di questa

declinazione della *causalità-2*, infatti, che le persone si innamorano, dimostrano il loro amore e riconoscono l’amore negli altri. Questo non rende l’amore meno vero, né il sentimento meno sincero o meno “autentico”. Che cosa causi quel sentimento sul livello più profondo della *causalità-1* non è assolutamente rilevante, perché quel livello, tagliando corto, non ha la minima idea di che cosa sia l’amore, perché lì non ci sono innamorati ma “soltanto” (ancora: nessuno voglia intendere questa parola come uno stigma di disvalore) comportamenti selezionati dall’evoluzione e, sul piano materiale dei nostri corpi, stringhe di istruzioni che liberano impulsi elettrochimici.

Qualcuno potrebbe essere scontento di questa spiegazione, perché a suo parere toglie all’amore, e ai sentimenti in genere, qualcosa che reputa importante. Ma chi la pensa così sbaglia, né più né meno, perché questa spiegazione, come ho mostrato, con quello che sul livello della mente simbolica chiamiamo “amore” non ha nulla a che vedere: non lo tocca, non lo modifica e non lo svisisce.

Consideriamo [...] il mito di Cupido, che svolazza in giro con le sue ali da cherubino facendo innamorare la gente con le frecce del suo piccolo arco. Questo è uno stereotipo quasi fumettistico, una tal caricatura che è difficile credere che qualcuno l’abbia mai presa alla lettera. Ma possiamo fare finta che, invece, sia stato proprio così: supponete che una volta ci fossero persone che davvero credevano che una freccia invisibile scoccata da un dio volante inoculasse un qualche virus che faceva innamorare la gente. E immaginate pure che un certo scienziato guastafeste, un giorno, si sia fatto avanti per dimostrare a tutti che la cosa era semplicemente falsa: non esiste alcun dio volante del genere. “Ha dimostrato che nessuno si innamora mai, che nessuno lo fa veramente. L’idea dell’innamoramento è solo una finzione carina – forse anche necessaria. Ma non capita mai”. Questo è ciò che avrebbe potuto dire qualcuno. Altri, si spera, avrebbero invece contestato una conclusione del genere: “No. L’amore è una cosa molto reale, come l’innamoramento. Solo che non è quello che la gente pensava che fosse. È altrettanto bello – forse anche di più. L’amore vero non ha bisogno di alcun dio volante”. [Dennett 2004: 297]

Le Tre Leggi che regolano il funzionamento del cervello di Dors Venabili non sono l’amore, ma non lo sono nemmeno la freccia di Cupido; né – che è quello che conta qui – le stringhe di istruzioni che viaggiano lungo le sinapsi. Le domande che Seldon pone a quella che sospetta essere un robot non hanno nessun senso, nello stesso modo in cui non avrebbe senso domandare a un essere umano se sia *davvero* innamorato o se la sua sensazione di innamoramento non sia piuttosto una terminologia convenzionale che serve per descrivere una complessa reazione chimica che avviene in esecuzione di una serie di istruzioni svincolate dalla sua volontà, il cui contenuto è stato stilato da milioni di anni di evoluzione, in maniera del tutto indipendente da *ogni* volontà. Non hanno senso perché, in definitiva, l’amore non è lì: non è nella freccia di Cupido, non è nei neuroni e non è nelle Tre Leggi, ma è soltanto nella mente simbolica. Quello che Seldon vuol sapere non è se il comportamento di Venabili – nel quale riconosce l’amore per lui – è un comportamento sincero nella testa di Venabili, ma se questo comportamento è prodotto da qualcosa di analogo a quello che produce in lui il sentimento d’amore verso di lei: Venabili è in grado di provare qualcosa di analogo a un sentimento? O meglio: Venabili è in grado di provare qualcosa di analogo a quello che, quando Seldon lo prova, definisce un sentimento?

A queste domande non ci sono risposte, e non perché Dors è un robot e Hari è un umano. Sarebbe così anche se entrambi fossero umani: io non so, e non ho modo di sapere, se il sentimento che dici di provare è analogo al mio. Quello che posso sapere è

soltanto *se io lo riconosco* come analogo al mio, se lo riconosco *sincero* nello stesso modo in cui sono pronto a definire sincero il mio. È il confine inesorabile del solipsismo: non so se il tuo sentimento è come il mio perché, in ultimo, non so se tu hai una mente come la mia, capace di provare sentimenti analoghi a quelli che provo io, perché, addirittura, non so né posso sapere se tu hai una mente *tout court*.

Il problema del solipsismo – e le sue varianti, come il dilemma del cervello nella vasca (Putnam 1981: 1-21) – non è un problema risolvibile, ed è per questo che abbiamo imparato a non considerarlo e che abbiamo adottato l'atteggiamento intenzionale. Alla fine, nessuno può escludere che ci sbagliamo nell'attribuire intenzionalità a determinati agenti esterni ("Il mio avversario, alla fine, era un programma"), ma sul lungo periodo e sulla larga scala le conseguenze negative derivanti dal singolo errore di classificazione sono praticamente annullate dal grande vantaggio che comporta il fatto che l'evoluzione abbia selezionato per noi delle convenienti scorciatoie.

Dors Venabili è umana o non lo è? Nel racconto di Asimov, a un certo punto c'è un tragico evento che pone fine alla sua esistenza. Dors viene attirata in una trappola, e senza accorgersene viene investita dall'energia di un dispositivo progettato per danneggiare i robot fino a distruggerli. Seldon la tiene tra le braccia, impotente mentre la guarda spegnersi e sapendo di non essere in grado di fermare l'inevitabile.

Dors scosse la testa adagio e sorrise ancor più debolmente.

"Addio, Hari, amore mio. Ricordati sempre... tutto quello che hai fatto per me."

"Non ho mai fatto nulla per te."

"Mi hai amata, e il tuo amore mi ha resa... umana."

I suoi occhi erano ancora aperti ma Dors aveva cessato di funzionare. [Asimov 1993: 259]

È probabile che Dors Venabili non sapesse dire se le sensazioni che provava con Seldon fossero equivalenti o analoghe a quello che Seldon definisce un sentimento di amore. L'amore è soltanto nella mente simbolica e non sappiamo se Venabili ne possedesse una. Sappiamo però che la possiede Seldon, e che Seldon ama Venabili, e che le Tre Leggi fanno sì che il comportamento di Seldon modifichi il comportamento di Venabili, e che il comportamento di Venabili diviene quello di un essere umano che si innamora di un'altra persona e che, alla fine, Dors è innamorata del suo Hari. Sì, è umana, e Hari può adesso piangere la morte della sua amata moglie.

Riferimenti bibliografici

Asimov, Isaac, [1985] 1988. *I robot e l'impero*. Milano: Mondadori.

---, [1988] 1991. *Preludio alla fondazione*. Milano: Mondadori.

---, 1993. *Fondazione anno zero*. Milano: Mondadori.

Brentano, Franz, [1867] 1989. *La psicologia di Aristotele*. Bologna: Pitagora.

Davies, Paul [2010] 2012. *Uno strano silenzio. Siamo soli nell'universo?* Torino: Codice.

Dawkins, Richard 1995. *Il fiume della vita. Che cosa è l'evoluzione?* Milano: Rizzoli.

---, [2006] 2007. *L'illusione di dio. Le ragioni per non credere*. Milano: Mondadori.

Dennett, Daniel Clement, [1987] 1993. *L'atteggiamento intenzionale*. Bologna: Il Mulino.

Persio Tincani, *"Sei umana?" Hari Seldon e l'atteggiamento intenzionale*

----, [1995] 1997. *L'idea pericolosa di Darwin. L'evoluzione e i significati della vita*. Torino: Bollati Boringhieri.

----, [2003] 2004. *L'evoluzione della libertà*. Milano: Cortina.

----, [2006] 2007. *Rompere l'incantesimo. La religione come fenomeno naturale*. Milano: Cortina.

de Duve, Christian, [2002] 2003. *Come evolve la vita. Dalle molecole alla mente simbolica*. Milano: Cortina.

Frank, Robert, 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: Norton.

Gillies, Donald, [1996] 1998. *Intelligenza artificiale e metodo scientifico*. Milano: Cortina.

Hume, David, [1748] 1996. *Ricerca sull'intelletto umano*. Roma-Bari: Laterza.

----, [1739] 2001. *Trattato sulla natura umana*. Milano: Bompiani.

Metzinger, Thomas, [2009] 2010. *Il tunnel dell'io. Scienza della mente e mito del soggetto*. Milano: Cortina.

Pinker, Steven, 1997. *How the Mind Works*. New York - London: Norton.

Putnam, Hillary, 1981. *Reason, truth and history*. Cambridge: Cambridge University Press.

Searle, John Ronald, 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences*, 3: 417-457.

Tincani, Persio, 2004. *Argomenti di giustizia distributiva*. Torino: Giappichelli.

Turing, Alan, 1950. "Computing Machinery and Intelligence." *Mind*, 59: 433-460.

----, [1969], 1994. *Intelligenza meccanica*. Torino: Bollati Boringhieri.