Pierpaolo Marrone

Macchine leibniziane?

Abstract: *Are robots capable of forming concepts? The problem can be addressed starting from a similar question, namely that relating to the possibility that animals are recognized as capable of forming concepts. Different positions have been addressed on this issue. A profitable possibility of analysis is offered by Davidson's skeptical position in his influential* Rational Animals*. I examine some of the objections made to the position that denies that animals can form concepts. I affirm that none of these are conclusive and I draw the idea that if the ability to form concepts must be recognized to animals, then the same position must also be supported with respect to robots.*

Keywords: *Concept, animal, robot, Davidson*.

L'onnipervasività delle macchine nelle nostre vite è un fenomeno che non credo abbia nemmeno bisogno di essere sottolineato per comprendere che solleva molti interrogative fondamentali per chi si occupa di filosofia[1]. Interrogativi forse ancora più esistenzialmente pressanti ci vengono dall'invasione prossima ventura dei robot nelle attività che costituiscono le nostre esistenze quotidiane[2]. Non mi sto riferendo alle piccole aspirapolveri che possono essere azionate da un app dedicata a distanza così come accade per condizionatori d'aria, sistemi di riscaldamento, lavatrici, dispositivi di sicurezza e così via. In fin dei conti possiamo interpretare questi sistemi meccanici come protesi della nostra volontà. Da questo punto di vista non sono molto interessanti, se non per il fatto che costituiscono un ulteriore capitolo della nostra mai conclusa ibridazione con le macchine e con gli artefatti, un vero e proprio universale umano presente in tutte le culture dove sia presente homo sapiens. Ciò che è più interessante dal mio punto di vista sono quei casi dove le macchine sono in grado di prendere decisioni autonomamente.

Occorre precisare che cosa significhi qui 'autonomamente'. Come è noto, il concetto di autonomia è alla base della nostra concezione dell'individuo come entità personale in grado di fare progetti e prendere decisioni che possono essere

1 R. Cingolani, *L'altra specie. Otto domande su noi e loro*, Bologna, il Mulino, 2019.

2 H. Fry, *Hello World: Being Human in the Age of Algorithms*, New York, Norton & Company, 2019.

Filosofia, anno LXV, 2020, pp. 63-74 • Mimesis Edizioni, Milano-Udine
Web: https://www.ojs.unito.it/index.php/filosofia • ISSN (print): 0015-1823, ISSN (online): 2704-8195 • DOI: 10.13135/2704-8195/5079

ricondotte all'individuo medesimo come al suo centro focale. Il soggetto autonomo sarebbe dunque quello che rintraccia in sé stesso la fonte delle sue decisioni. Questo concetto, che è alla base tanto della concezione cartesiana della sostanza quanto dell'ossessione della filosofia moderna e contemporanea per il soggetto, è altamente controverso[3]. Se noi intendiamo l'autonomia come una forma di spontaneità, allora sembrerebbe che siamo costretti ad entrare in inevitabili paralogismi. Dal momento che noi non siamo né enti che si sono autocreati né enti in grado di autoprogrammarsi, è difficile credere che le nostre decisioni siano autocreate e autoprogrammate. Eppure il concetto di autonomia ha a che fare con l'autoprogrammazione, dal momento che significa precisamente la capacità di dare norme di condotta a sé stessi. Quindi, quando si parla della nostra capacità di prendere decisioni si intende qualcosa di molto complicato e non la semplice idea che l'azione è la manifestazione visibile di una volontà che non ha altro fondamento se non in sé stessa.

Non entro nel merito della questione relativa alla libertà del volere, che inevitabilmente sorge ogni volta che si discute di autonomia. Non è mia intenzione né affermare né negare tale libertà. Quello che mi è sufficiente stabilire è che la nostra autonomia è condizionata da molti fattori che vanno dal nostro codice genetico, all'ambiente nel quale ci è toccata la sorte di essere cresciuti ed educati, all'intreccio tra la nostra biologia individuale e le condizioni socio-ambientali che troviamo costituite attorno a noi e che noi non abbiamo in alcun modo formato. È dentro a questo format, appunto, che si esercita la nostra capacità di scelta. Noi elaboriamo le informazioni che ci giungono da una selezione di stimoli ambientali e di stimoli interni, selezioniamo quelle che crediamo essere pertinenti, attingiamo dalla nostra esperienza passata quanto ci può essere utile per rintracciare degli schemi persistenti, adottiamo, insomma, assieme alle informazioni pregresse tutta una serie di strategie condizionali del genere 'se… allora'. È con queste premesse che si esercita la nostra azione, dove l'autonomia deve essere considerata una etichetta che individua una classe di azioni che facilmente noi riconduciamo a noi stessi come al loro punto di partenza contestuale[4].

Le azioni si esercitano spesso sulla base della nostra esperienza passata e questa esperienza pregressa ha valore per noi soprattutto perché vi ricerchiamo e spesso rintracciamo degli schemi persistenti. Alcuni di questi schemi persistenti noi li chiamiamo 'concetti'. I concetti sono strumenti indispensabili al nostro orientamento nel mondo. Ci fanno capire che cosa fare ('non avvicinarti troppo a quello che sembra solo vagamente un gatto un po' sovradimensionato'), ci fanno comprendere quali aspettative è razionale avere ('non è la stessa cosa assumere una sostanza psicotropa e una minestra di asparagi'), ci permettono di fare piani per il futuro (ad esempio facendoci comprendere che una vacanza è una cosa diversa da

3 G. Dworkin, *The Theory and Practice of Autonomy*, Cambridge, Cambridge University Press, 1988; B. Rosslenbroich, *On the Origin of Autonomy*, New York, Springer, 2014; J. Schneewind, *The Invention of Autonomy: A History of Modern Moral Philosophy*, Cambridge, Cambridge University Press, 1997.

4 R. Sapolsky, *Behave: The Biology of Humans at Our Best and Worst*, New York, Vintage, 2018.

un trasloco). Tutti questi esempi si basano su categorizzazioni di esperienze passate nostre o altrui e rientrano nella categoria dei concetti empirici o a posteriori, che necessitano dell'esperienza per essere formati e utilizzati[5].

Sembrerebbe che la capacità di formare concetti sia un presupposto dell'azione razionale. Quale che sia la definizione di razionalità che vogliamo utilizzare, la capacità di generalizzare le esperienze al fine di orientarsi nel futuro e di interpretare correttamente il passato ne è una componente essenziale. Questa componente sembra essere particolarmente necessaria a una razionalità concepita strumentalmente, ossia intesa come la capacità di scegliere i mezzi adeguati in vista del raggiungimento di fini determinati. Sembrerebbe che gli animali siano effettivamente capaci di razionalità strumentale[6]. Le prove sono le sorprendenti capacità adattive degli organismi viventi e la loro sopravvivenza nel corso del processo evolutivo. Ma possedere ed esibire comportamenti che possono essere interpretati secondo i criteri della razionalità strumentale non sembra essere equivalente alla capacità di formare concetti. Certe piante hanno sviluppato meccanismi adattivi per attirare insetti da utilizzare come impollinatori, ma pare azzardato dire che questo abbia a che fare con la capacità delle piante di formare dei concetti. Quindi si potrebbe inferirne che una cosa è definire un comportamento come razionale e un'altra cosa definire un agente come razionale. Ci possono essere comportamenti che sono interpretabili secondo una qualche definizione di razionalità (ad esempio quella strumentale), ma non definiscono affatto l'agente che ha quel comportamento come razionale. Infatti, se estendiamo eccessivamente il concetto di razionalità di un'azione secondo la definizione di razionalità strumentale lo faremo coincidere con quello di ragione sufficiente ('dal momento che tutto quello che accade, accade per una ragione, allora tutto accade come dovrebbe accadere'), ma questo lo renderebbe inutilizzabile per comprendere che cosa significa essere un agente razionale.

Si può quindi agire razionalmente senza essere degli agenti razionali in senso pieno. Questo solleva immediatamente il problema di che cosa renda razionale un agente[7]. La prima risposta che forse intuitivamente a molti verrebbe in mente è che l'agente deve esibire una qualche forma di intenzionalità[8]. Come possa essere accertata questa forma è una questione controversa. Forse negli animali un indice

5 G. Löhr, *Embodied Cognition and Abstract Concepts: Do Concept Empiricists Leave Anything Out*, in "Philosophical Psychology", XXXII (2019), n. 2, pp. 161-185; O. Vasilecas, D. Bugaite, J. Trinkunas, *Knowledge Expressed By Ontology Transformation into Conceptual Model*, in "Communication and Cognition", XL (2007), n. 1-2, pp. 13-24.

6 K. Andrews, *The Animal Mind: The Philosophy of Animal Cognition*, London, Routledge, 2015; H. J. Glock, *Animals, thoughts, and concepts*, in "Synthese", CXXIII (2000), n. 1, pp. 35–64; H. J. Glock, *Can animals act for reasons?*, in "Inquiry", LII (2009), n. 3, pp. 232-254; K. Oslo, *Rationality in the Domesticated Dog and Other Non-Human Animals*, in "Teorema", XXIX (2010), n. 2, pp. 135-145.

7 M. Jago, *The Problem of Rational Knowledge*, in "Erkenntnis", LXXIX (2014), n. 6, pp. 1151-1168; L. Roelofs, *Rational Agency without Self-Knowledge: Could 'We' Replace 'I'?*, in "Dialectica", LXXI (2017), n. 1, pp. 3-33.

8 D. Davidson, *Essays on Events and Actions*, Oxford, Clarendon Press, 1980; D. Dennett, *The Intentional Stance*, New York, Bradford, 1989.

indiretto può essere considerato la capacità di mentire. Se riesco a ingannare un altro animale, allora questo è indice della presenza di un pensiero complesso, qualcosa che potrebbe essere schematizzato come 'desiderare che qualcuno desideri qualcosa e indurre un comportamento affinché lo faccia'. Quando la mia gatta, miagolando e dirigendosi verso il poggiolo dove sta la ciotola con il suo cibo, mi induce ad alzarmi dalla mia poltrona sulla quale poi lei va a sedersi, penso che lei sia riuscita a farmi fare qualcosa che non avrei volute fare per raggiungere con maggiore facilità un suo obiettivo. Se riesce a farlo più di una volta posso cominciare a pensare che questo non sia accaduto per caso e che la mia gatta abbia una capacità di progettare le azioni future con una certa consapevolezza, grazie all'adozione di schemi ricorrenti del genere 'se miagolo insistentemente dirigendomi verso il poggiolo, allora lui si alzerà dalla mia poltrona preferita e la libererà'. Naturalmente questa è una schematizzazione che potrebbe essere del tutto impropria. Non è necessario che la mia gatta abbia il concetto di 'poggiolo' o di 'poltrona preferita' o di 'insistentemente'. Tuttavia, qualcosa del genere deve pur accadere nel suo pensiero affinché lei effettivamente riesca a progettare di farmi alzare dalla mia poltrona preferita che è anche la sua[9]. Ora, il problema è che cosa precisamente deve accadere.

Noi attribuiamo con una certa facilità l'atteggiamento intenzionale per spiegare molte attività legate soprattutto alle azioni degli organismi viventi. Ne facciamo un uso inflazionistico in effetti. Questo uso inflazionistico dovrebbe secondo Davidson essere uno degli elementi che ci dovrebbe consigliare prudenza e sospetto. Forse stiamo maneggiando concetti senza essere ben consapevoli delle loro implicazioni e forse stiamo dando per scontato che pensiero, intenzionalità, razionalità siano implicati in un qualche insieme olistico che accomuna per lo meno una gran parte degli organismi viventi. Ma come noi riconosciamo normalmente quelle caratteristiche che pensiamo essere razionali, tra le quali l'attribuzione di intenzionalità e la capacità di formare schemi e di riconoscere schemi (quanto normalmente chiamiamo produrre concetti)? Noi riconosciamo negli altri – crediamo di riconoscere in altri – un'attitudine che è in noi, che Davidson nel suo influente articolo *Rational Animals* chiama "atteggiamento proposizionale"[10], ossia la capacità di avere cose come una credenza, un desiderio, un'intenzione, la capacità di vergognarsi, di ingannare, di distinguere il vero dal falso.

L'atteggiamento intenzionale, la vergogna, la menzogna, la distinzione tra vero e falso possono essere tutti raggruppati sotto il concetto di credenza. Tuttavia, come si può essere sicuri che un qualche organismo o un qualche sistema con un'organizzazione cognitiva di una certa complessità abbiano una credenza? Per Davidson esiste una sorta di test per passare l'esame dell'ascrizione di una credenza: si deve essere capaci di formare il concetto di credenza. Questa è per Davidson un'intuizione centrale e della massima importanza. Nessun animale può avere una creden-

9 K. Oliver, *Duplicity Makes the Man Or, Can Animals Lie?*, in *The Philosophy of Deception*, a cura di C. Martin, Oxford, Oxford University Press, 2009, pp. 104-117.

10 D. Davidson, *Rational Animals*, in *Actions and Events*, a cura di E. LePore, B. McLaughlin, Oxford, Blackwell, 1985, pp. 473-480.

za se non ha anche il concetto di credenza. Per Davidson questo non è possibile e per una ragione diretta. Nessuno può avere una credenza a meno che non comprenda anche l'eventualità di essere ingannato o di essere in errore. Questo significa possedere le necessarie qualità mentali per distinguere il vero dal falso, ossia la capacità di discriminare tra credenze vere e credenze false. Non so se si può dire che Davidson ha presentato un argomento per provare questa sua intuizione. A me pare di no, ma non credo nemmeno che questo sia del tutto rilevante. Io credo che Davidson abbia prodotto qualcosa di diverso in questa sua posizione, ossia abbia introdotto una rete sofisticata di concetti e problemi, per mostrare come noi non siamo in possesso di idee sufficientemente chiare su tutta la questione, anche se una delle sue conclusioni più forti è che per avere una credenza è necessario avere il concetto di credenza e questo lo porta a sostenere che non è affatto chiaro in che senso un animale possa avere una credenza. Questa conclusione qualifica, almeno *prima facie*, la sua posizione come fortemente scettica a proposito dell'ascrizione di capacità concettuali negli animali.

Pensiamo a questo esempio: vedo un cane che insegue un gatto che corre in una piazza che ha un solo grande albero, una maestosa magnolia, poniamo. Questa maestosa magnolia è anche l'albero più antico di tutto il quartiere. Il gatto gira attorno alla magnolia e si rifugia sull'albero. Il cane abbaia furiosamente verso l'albero. In che senso il cane ha la credenza che il gatto sia sulla magnolia? Si è formato il concetto di gatto, si è formato il concetto di albero? Nessuno di noi penserebbe che si sia formato il concetto di albero più antico di tutto il quartiere. Non è certamente in grado di sostituire 'magnolia' a 'albero più antico di tutto il quartiere' nei contesti pertinenti *salva veritate*. Ma è davvero necessario che accada qualcosa del genere[11]? Formare un concetto può essere considerata un'operazione che comporta diversi gradi di complessità. Per quanto ne sappiamo una delle complessità maggiori è permessa dalla formazione di concetti all'interno del linguaggio umano. In questo ambito è sensato richiedere che sia possibile sostituire 'magnolia' a 'albero più antico di tutto il quartiere' *salva veritate*, ai fini di una verifica di una corretta comprensione della credenza che la magnolia è l'albero più antico di tutto il quartiere. Se un parlante normale non è in grado di fare le usuali sostituzioni nei contesti rilevanti *salva veritate*, allora concludiamo che non ha formato correttamente i concetti. Tuttavia, quanto vale correttamente per i parlanti umani, dal momento che non può essere richiesto al cane, deve farci concludere che il cane non si è formato il concetto di gatto o il concetto di albero? In una qualche maniera se la magnolia fosse circondata da una aiuola di fiori il cane dovrebbe sapere che il gatto non può essersi arrampicato sui fiori[12]. Deve aver pensato in qualche modo che dal momento che il gatto è un animale che si arrampica agilmente su superfici tendenzialmente verticali, allora deve essersi arrampicato sull'albero e non sui fiori. Anche se non ha il concetto di credenza nel senso richiesto da Davidson il cane

11 J. Toribio, *The Animal Concepts Debate: A Metaphilosophical Take*, in "Teorema", XXIX (2010), n. 2, pp. 11-24.

12 A. Colin, *Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach*, in "Erkenntnis", L (1999), n. 1, pp. 33-40.

deve pur sempre avere la credenza che il gatto sia sull'albero. Deve avere fondati motivi a sostegno di questa sua credenza, perché altrimenti noi ci troveremmo in difficoltà a spiegare perché il cane continui ad abbaiare all'albero. Questa idea dà in qualche modo per acquisita l'idea che il pensiero per quanto nell'homo sapiens strettamente intrecciato al linguaggio e per quanto ci sia impossibile stabilire nettamente i confini tra pensiero e linguaggio (così come tra credenza e atteggiamento proposizionale), tuttavia è presente anche in organismi che non possiedono il linguaggio verbale, mentre possiedono altri sistemi di comunicazione anche interspecifici. Non sembra esserci maniera più economica e migliore di spiegare il comportamento del cane se non attraverso la sua rete di credenze che comporta anche la formazione pregressa di reti concettuali[13]. Quanto vale per il cane vale ovviamente anche per il gatto che quando il cane lo insegue 'sa che il cane sa' che non deve permettergli di raggiungere la maestosa magnolia al centro della piazza. È vero, tuttavia, che questo è un esempio di come molto facilmente siamo indotti a interpretazioni inflazionistiche nell'attribuzione di reti concettuali, credenze, intenzionalità? Non credo, mentre se non sapessi con una certa quantità di particolari significativi che cosa è un missile, potrei pensare che il missile che ha colpito un bersaglio abbia manifestato l'intenzionalità di colpire quel bersaglio dopo averlo individuato, averlo discriminato dal resto dell'ambiente, essersi formato delle reti concettuali che gli hanno permesso di eseguire il compito che aveva scelto di compiere. Ovviamente, questa sarebbe una spiegazione completamente sbagliata e chi mi facesse notare che dietro al missile che colpisce il bersaglio ci sono degli operatori umani che prendono delle decisioni non fornirebbe semplicemente una spiegazione alternativa alla mia, bensì sarebbe in possesso di una spiegazione migliore. La sua rete esplicativa è superiore alla mia perché descrive la realtà in una maniera migliore della mia, anche se entrambi possediamo un concetto di credenza che ci distingue dal cane e dal gatto dell'esempio precedente.

Quello che c'è di comune ai nostri differenti e elaborati sistemi di credenze, è che possono essere descritti come veri o falsi. Questo sembra non poter accadere per Davidson per altri sistemi come quelli che vengono implementati dagli animali oppure dai neonati, oppure dai malati di Alzheimer. Avere un concetto significa avere la credenza che lo schema mentale che lo definisce sia correlato all'abilità di distinguere alcune cose da altre in un numero significativo di contesti qualificati. Questa è una richiesta eccessiva? Naturalmente noi abbiamo solo indici indiretti della capacità di altre menti di formare dei concetti in questo senso ed eventualmente di avere il concetto di credenza. Ma che cosa significa avere una credenza? Penso che possedere una credenza possa essere descritto come pensare che qualcosa nel mondo è o è stato o sarà così e così. Ascrivere una credenza a qualcuno è ritenere che abbia una strategia di comportamento nel mondo basata su queste credenze[14]. Per avere una credenza in questo senso è anche necessario saper spe-

13 P. Carruthers, *Animal Minds Are Real, (Distinctively) Human Mind Are Not*, in "American Philosophical Quarterly", L (2013), n. 3, pp. 233-248.

14 N. Alechina, B. Logan, *Belief ascription under bounded resources*, in "Synthese", CLXXIII (2010), n. 1, pp. 179-197.

cificare che questa è una credenza? Io penso di no. È certamente necessario avere quella che può essere chiamata credenza relativa alla credenza se facciamo delle operazioni di pensiero riflessive[15]. È ovviamente vero che esistono sistemi biologici che possiedono credenze e concetti nel senso di Davidson, ad esempio homo sapiens è un sistema di questo genere. Mi pare anche evidente che l'ascrizione di intenzionalità sia tanto un'operazione cruciale di descrizione esaustiva di un'azione quanto una strategia naturale.

Il problema non è quindi l'intenzionalità per noi o la capacità di avere una credenza e di credere di credere, quanto il fatto che questa non sia una condizione eccessivamente esigente per interpretare altri sistemi di pensiero. Animali che non sono in grado di avanzare pretese di riflessività sui propri sistemi di pensiero non sono nemmeno in grado di avere una credenza in senso forte? Probabilmente è così, e se è corretto suggerire che è questa mancanza a doverci far pensare che una cosa è la capacità di discriminare e un'altra cosa, ossia un'operazione qualitativamente diversa, è avere un atteggiamento intenzionale, non deve farci concludere che è necessario avere un'intenzionalità in senso riflessivo (precisamente il senso che suggerisce Davidson) per poter dire che si ha una credenza, differenziata dalla semplice capacità discriminatoria che è propria dei gatti, dei cani e di altri animali. Se poi confrontiamo esseri viventi e sistemi artificiali, una delle prime distinzioni che si potrebbero fare è che i sistemi viventi ci costringono, quando interpretiamo i loro comportamenti, a fare propria un'assunzione di razionalità che nei sistemi artificiali non può essere presente se non come un'attribuzione del progettista.

Ciò che Davidson afferma, secondo me, non è precisamente la negazione dell'intenzionalità a sistemi meno complessi di quello umano adulto e con prestazioni cognitive medie, quanto piuttosto la nostra difficoltà a segnare precisamente dove inizi l'incapacità di essere intenzionali e di avere una credenza. È molto probabile che né il cane né il gatto che è inseguito dal cane possiedano il concetto di verità, dal momento che il vero e il falso non sono nelle cose, ma nelle nostre descrizioni di stati del mondo e per quanto ne sappiamo non possiamo dire che il cane e il gatto dispongano delle descrizioni proposizionali del mondo. Tuttavia, questo non significa che non abbiano delle descrizioni del mondo come appare a loro. Questa è stata la posizione ad esempio di Stitch, il quale pensava che se gli animali hanno credenze, allora la migliore spiegazione del loro comportamento deve essere nei termini di una teoria esplicativa analoga a quella che adottiamo per il comportamento umano[16]. Davidson probabilmente riterrebbe questa posizione come un'indebita assunzione del punto che si tratta di dimostrare, ma il problema secondo me è precisamente se sia possibile una dimostrazione di qualcosa del genere. Dal momento che ci dobbiamo affidare ad indici comportamentali non sarà mai possibile, se non attraverso questi indici indiretti, avere credenze vere giustificate sulle credenze che determinati sistemi

15 J. E. Burgos, *About Aboutness: Thoughts on Intentional Behaviorism*, in "Behavior and Philosophy", XXXV (2007), n. 1, pp. 65-76.

16 S. P. Stitch, *Do Animals Have Beliefs*, in "Australasian Journal of Philosophy", LVII (1979), n. 1, pp. 15-28.

hanno o potrebbero avere. Si potrebbe pensare alla credenza come a un sistema complesso che ha dei gradi e che si colloca lungo un continuum. Forse anche il pensiero è un continuum. Ciò non è per nulla in contraddizione con il fatto che alcune prestazioni cognitive sono state possibili unicamente a partire dalla comparsa del linguaggio e dei meccanismi di anticipazione, previsione, rappresentazione, retroazione che il linguaggio permette e grazie ai quali ha costituito uno straordinario vantaggio evolutivo per homo sapiens.

La condizione di credere di credere per avere la certezza che si abbia l'intenzionalità è, tuttavia, una condizione troppo esigente. Di più: penso che anche la condizione di essere capaci di avere un atteggiamento proposizionale come condizione necessaria per avere la capacità di formare concetti sia troppo esigente. In realtà, io penso che sia troppo esigente anche rispetto alla condizione di verità. Naturalmente né il cane né il gatto sono in grado di formare proposizioni, ma se accettiamo che il cane sappia che sta rincorrendo un gatto e pensiamo che sappia discriminare tra un gatto e un oggetto meccanico telecomandato, allora potrà anche sapere che l'oggetto telecomandato non è un gatto, sebbene sulle prime sia stato ingannato e indotto a inseguirlo. Aveva una credenza sull'oggetto telecomandato che si è dimostrata inadeguata alle sue aspettative, se non vogliamo usare la parola 'falsa', e questa credenza inadeguata lo ha costretto a resettare i suoi comportamenti rispetto all'oggetto telecomandato e rispetto a molte altre cose che sono nel suo ambiente circostante. Si potrà allora dire che il cane è stato capace di distinguere tra una credenza inadeguata e una adeguata. Del resto, I fatti non ci stanno segnalando proprio questo? Se siamo disposti ad ammetterlo, non dovremmo essere anche disposti a concedere che vero e falso sono proprietà delle credenze prima ancora di essere proprietà degli enunciati? Del resto non sarebbe proprio un vantaggio evolutivo essere capaci di discriminare credenze vere (adeguate) da credenze false (inadeguate)?

Prima ancora di tutto questo deve situarsi la capacità di formare concetti. Quali potrebbero essere alcune delle caratteristiche che è necessario esibire per passare un test sulla capacità di formare concetti? Un test potrebbe essere la capacità di migliorare le proprie prestazioni a fronte di errori passati. Le prove che molte specie animali sono in grado di un apprendimento del genere abbondano[17]. Ritengo che queste prove semplicemente rendano giustificata l'idea che gli animali formano concetti.

Penso che la capacità di formare concetti proprio nel senso di formare schemi utilizzabili ai fini della discriminazione nell'ambiente e ai fini dell'autoapprendimento sia una condizione necessaria del pensiero, ma credo anche che si tratti di una condizione assolutamente non sufficiente. Infatti l'autoapprendimento e la capacità di discriminare sono caratteristiche anche di sistemi esperti artificiali da tempo e non per questo riteniamo che questi sistemi pensino. Si pensi ai progressi che Google Translate ha mostrato negli ultimi anni, passando da un

17 C. Safina, *Beyond Words: What Animals Think and Feel*, London, Picador, 2016; P. Godfrey-Smith, *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*, New York, Farrar Straus & Giroux, 2017.

sistema generalmente rozzo a un sistema capace in molti casi di performance di traduzione linguistica notevoli. Tutto questo è accaduto grazie alla enorme potenza di calcolo che accedendo agli archivi di big data può in millisecondi confrontare le occorrenze di una parola oppure di un'espressione e scegliere il contesto più appropriato per quell'espressione. Nel settore della sicurezza le tecniche di riconoscimento facciale hanno mostrato impressionanti capacità di discriminare volti in contesti caratterizzati dalla presenza di grandi folle, come ad esempio negli stadi. I sistemi esperti che finalizzano l'apprendimento automatico, il *deep learning* e il *machine learning*, non vivono più in un circoscritto mondo deterministico quale poteva essere considerato la scacchiera sulla quale Deep Blue sconfisse Garry Kasparov nel 1997, bensì in un mondo dove sono le macchine a fare esperienza[18]. Nel nostro mondo noi siamo soggetti che conoscono in maniera intuitiva molte più cose di quante riusciamo mai a comunicare. Ad esempio come si scende da un'automobile (un compito ancora difficilmente eseguibile da un robot) o riconoscere un volto. I sistemi esperti elaborano schemi automatizzati dall'analogo dell'esperienza e questo analogo sono i big data. Mediante l'utilizzo di funzioni statistiche che classificano le ripetizioni, i dati vengono interpretati come probabilità di eventi e la probabilità di un evento è una classificazione con capacità predittiva. Così una macchina è in grado di riconoscere con una data probabilità se una voce umana appartiene a un uomo o a una donna, se si tratta di soggetti adulti, anziani o di bambini. L'algoritmo che riconosce le immagini opera in base ad occorrenze statistiche relative alla posizione di un oggetto, alla sua illuminazione, alla sua angolatura rispetto al piano e allo sfondo. Tutte queste sono stilizzazioni, si dirà, ma ai fini dell'apprendimento in numerose scienze, ad esempio l'anatomia e l'ornitologia, sono estremamente efficaci i disegni che selezionano i tratti che devono essere ritenuti rilevanti in quella conoscenza. L'occorrenza del singolo uccello o del singolo organo che ci troviamo sotto gli occhi sono riconosciuti nel *deep learning* per stratificazioni di astrazioni – questo è il significato di *deep*, che indica non tanto una profondità dietro la realtà, quanto piuttosto un sistema reticolare di astrazioni a molteplici livelli[19] –. Così un volto viene riconosciuto a partire dall'analisi dei singoli pixel sino a riuscire a contornare quello che appare con alta probabilità come una faccia e così via. Non diversamente un sistema di guida autonomo discrimina altri veicoli da pedoni che stanno attraversando la strada o un algoritmo diagnostico contorna le cellule tumorali in un'immagine diagnostica. Il punto basilare, che è il fondamento di apprendimenti sempre più estesi (il cosiddetto *deep neural network*), è sempre l'individuazione di schemi ricorrenti (che permettono anche di riconoscere anomalie statistiche, come transazioni fraudolente con carte di credito, ad esempio). In questo senso mi pare indubitabile sostenere che le macchine formano concetti, come formano concetti anche gli animali.

18 A. Vespignani, *L'algoritmo e l'oracolo. Come la scienza predice il futuro*, Milano, il Saggiatore, 2019.

19 C. Accoto, *Il mondo ex machina. Cinque brevi lezioni di filosofia dell'automazione*, Roma, Egea, 2019.

Si può poi dire un'altra cosa. L'interpretazione bayesiana delle probabilità interpreta le frequenze come gradi di fiducia e di approssimazione alla verità. È possibile dare un'interpretazione simile del *deep learning*, come un incremento del grado di approssimazione al miglior risultato per l'algoritmo che si approssima al vero? Se è così, allora sembra difficile non stabilire una qualche parentela ulteriore tra noi, gli animali non umani e loro, le macchine che abbiamo creato. Anche le macchine oltre a formare concetti si avvicinano al vero. Non ha molto senso, io credo, affermare che siamo noi che abbiamo elaborato l'algoritmo iniziale e quindi le macchine che implementano le prestazioni che sono consentite da questi algoritmi non formano realmente concetti. Lo stesso si potrebbe dire della nostra discendenza biologica, ma questo ovviamente non avrebbe alcun senso. C'è però una cosa in comune tra sistemi di *deep learning* e apprendimento automatico e la nostra discendenza biologica: spesso noi non sappiamo come gli output vengono prodotti. Da questo tema si è originato tutto un filone di ricerca che sottolinea le speranze e i rischi potenziali della AI[20]. L'idea molto spesso sottostante a questi timori è che possiamo affidarci alle macchine solo se siamo in grado di comprendere nel dettaglio il loro funzionamento.

Un concetto può essere considerato un algoritmo, ossia una procedura per produrre un risultato (una schematizzazione, una conoscenza, un raggruppamento insiemistico) in un numero finito di mosse. Il trend che stiamo vivendo è quello dove una quantità impressionante di decisioni sono prese da algoritmi non tanto in attività di scarso interesse perché altamente ripetitive, quanto in processi decisionali che hanno a che fare con le pratiche che noi associamo alla democrazia, pratiche che vanno dalla dimensione burocratica a quella legislativa a quella legale. Esistono algoritmi che selezionano le persone alle quali controllare i bagagli all'aeroporto, algoritmi che prevedono il tasso di evasione fiscale di determinate categorie, algoritmi che ci consigliano i potenziali partner da contattare su siti di *dating*, algoritmi di previsione delle nostre preferenze elettorali disegnate in base alle nostre attività on line.

Oramai, numerosi studiosi incominciano a parlare dei pericoli dell'algocrazia[21], termine che indica non tanto un mondo governato da computer, quel mondo propagandato da innumerevoli film di fantascienza, quanto il nostro mondo e la tendenza ad affidarci alle macchine per prendere decisioni di rilevanza pubblica. In questa preoccupazione si condensano due problemi diversi, entrambi relativi a una questione di opacità. Da una parte c'è il problema dell'opacità nell'utilizzo dei nostri dati, quei dati che noi produciamo continuamente. Non è quasi mai chiaro come questi dati vengono raccolti e in che modo vengano utilizzati. Ovviamente vengono utilizzati a scopi di profilazione commerciale, ma possono essere usati anche per una propaganda politica mirata a gruppi ristretti di elettori. Si sono

20 P. Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, London, Penguin, 2017; N. Bostrom, *Superintelligence. Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2016.

21 J. Danaher, *The Threat of Algocracy: Reality, Resistance and Accommodation*, in "Philosophy & Technology", XXIX (2016), n. 3, pp. 245-268.

paragonati questi dati al nuovo petrolio ed in effetti c'è stato qualcuno che ha sostenuto che le big company che manipolano e gestiscono queste enormi masse di informazioni dovrebbero pagare per ottenerle (però c'è un problema: noi le abbiamo sì volontariamente consegnate, ma non in cambio di niente, bensì in cambio di socialità). Chi ha accesso a questa enorme massa di dati, non sappiamo come vorrà utilizzarla, dunque. Ma questa è solo la prima preoccupazione. Ce n'è un'altra relativa non all'opacità delle decisioni, ma a quella che potremmo chiamare opacità epistemica. La potenza di calcolo di questi sistemi, la loro capacità di autoapprendere e di autocorreggersi è tale che cominciano a manifestarsi casi nei quali non comprendiamo come si sia giunti al risultato finale. Alcuni temono che questo produrrà ben presto generazioni di algoritmi che sorpasseranno di molto i limiti cognitivi degli esseri umani relativamente non alla produzione del risultato, come già accade, bensì piuttosto alla comprensione del risultato stesso. Questo pone questioni, sulle quali ora non entro, che riguardano l'estensione dell'autorità politica e la legittimità delle decisioni, ossia la nostra capacità di partecipare in maniera informata alle decisioni politiche. La sfida della opacità sistemica del *deep learning* e dell'apprendimento automatico è probabilmente solo all'inizio, perché non mancano pericoli di discriminazione selettiva, che potrebbero essere considerati dei pregiudizi implementati dalla macchina[22]. E che cosa è un pregiudizio se non uno schema concettuale non rappresentativo o solo parzialmente rappresentativo della realtà empirica? Le macchine sono in grado di produrre concetti anche se non sono ancora in grado di pensare e quindi certamente non sono in grado di produrre una credenza sulla propria produzione concettuale.

Le indicazioni di Davidson sul problema dell'ascrizione di capacità concettuali come devono essere prese? Devono essere considerate come una posizione risolutamente scettica sulla capacità degli animali di avere concetti, come alcuni le hanno interpretate, incoraggiati in questo da Davidson stesso? O devono essere considerate come un'indicazione della nostra difficoltà di ascrivere delle credenze? Oppure ancora devono essere prese per l'affermazione olistica di una determinata concezione del pensiero che ricorda quella di Leibniz per il quale tutte le unità spirituali, che lui chiamava monadi e che esauriscono il dominio del reale, hanno un'attività mentale, differendo per gradi, dalla percezione oscura a quella consapevole, che lui chiamava appercezione e che in Kant diverrà il centro focale del pensiero del soggetto trascendentale? Tra queste tre opzioni io credo che la prima debba essere esclusa, mentre la seconda e la terza devono essere accolte. Si tratta di due opzioni che sono profondamente solidali tra di loro, perché ipotizzano un continuum tanto nella credenza quanto nel pensiero. In altre parole, la credenza non è affatto una qualità che è o presente o è assente, ma assomiglia a un colore che può essere più o meno intenso. Quanto più i sistemi cognitivi divengono complessi nella scala evolutiva, tanto più si ha una credenza maggiormente 'satura'. Ma è molto difficile dire dove la credenza precisamente compare e dove scompare. Così la mia idea è che la

22 F. Bacchini, L. Lorusso, *Race, again. How face recognition technology reinforces racial discrimination*, in "Journal of Information, Communication and Ethics in Society", XVII (2019), n. 2, pp. 321-335.

capacità di formare concetti sia una condizione del tutto necessaria affinché ci sia pensiero, ma è una condizione assolutamente non sufficiente. Molti dei dubbi che vengono sollevati sulla capacità delle macchine a produrre e utilizzare dei concetti, oramai anche in maniera autonoma, come accade con i software di riconoscimento facciale, derivano da problemi che riguardano l'opacità dei processi di apprendimento. Questa opacità genera preoccupazioni quando diviene opacità nei processi decisionali, come si è detto per quanto riguarda l'algocrazia, perché ci espropria di opportunità decisionali. Se questo sia sempre un male è argomento controverso. Così come è naturalmente controverso se le macchine che noi costruiamo saranno mai capaci un giorno di pensare. Non era questo l'argomento di queste pagine che riguardavano invece la capacità delle macchine di produrre concetti. La mia tesi è stata che se questa capacità viene riconosciuta a determinati animali, allora non vi è motivo di non riconoscerla a determinate macchine. Non basta questo affinché ci sia pensiero e credenza nella misura in cui pensiero e credenza sussistono nell'essere umano. Ma le basi sono state gettate dall'evoluzione in un caso e dalla nostra creatività nell'altro. Che questo dia sostanza all'haiku che Adam, l'androide del romanzo di Ian MacIwan, *Machines like me*, compone alla fine della sua esistenza ("Our leaves are falling / Come spring we will renew/ But you, alas, fall once"[23]) nessuno è in grado ora di dire, ma tutti noi siamo in grado di immaginare.

23 I. McEwan, *Machines like me*, London, Penguin, 2019, p. 280.