

## di Gianfranco Pecchinenda

In uno dei suoi tanti capolavori - *L'eterno marito* - Fëdor Dostoevskij così descrive lo stato d'animo del protagonista principale della storia: "Pavel Pavlovič voleva ucciderlo, ma l'idea di ucciderlo non gli era mai venuta in mente. Più brevemente: Pavel Pavlovič voleva uccidere, ma *non sapeva* di voler uccidere".

Siamo nel 1871, e il genio che meglio di ogni altro ha saputo scandagliare i dilemmi dell'io e le dinamiche oscure del rapporto tra l'individuo e se stesso, configurando un'originale filosofia esistenzialista, ci mette in guardia sulla sottile quanto determinante differenza tra il "volere", e il "sapere di volere", ovvero quello stato di consapevolezza della propria intenzionalità, da parte dell'attore stesso, cui non necessariamente tutti sono in grado di pervenire.

Dal suo canto un altro grande scrittore russo, Vladimir Nabokov, sosteneva che "se oltre a sapere che *io sono* so anche di saperlo, allora appartengo alla specie umana".

Ed è proprio all'analisi di questo fenomeno così caratterizzante per la nostra specie, che il neuroscienziato Stephen M. Fleming, tra i massimi esperti al mondo di *metacognizione* e direttore del prestigioso *MetaLab* allo University College di Londra, dedica il suo recente lavoro dal titolo *Conoscere se stessi. La nuova scienza dell'autoconsapevolezza* (Raffaello Cortina, 2022).

La *metacognizione* è la capacità che gli esseri umani possiedono di "riflettere sul proprio stesso pensiero": dal greco antico *metá*, che significa "dopo" o "oltre". Si tratta di una caratteristica fragile, che da secoli affascina pensatori e filosofi di ogni dove.

Nel celebre trattato di Linneo - il *Systema Naturae* - essa viene definita in una sola, semplice formula: *Nosce te ipsum*, "Conosci te stesso". John Locke, nel secondo libro del suo *Saggio sull'intelletto umano* (1690), aveva già definito *senso interno* questa capacità di percepire interiormente le operazioni che compie la nostra mente: "Sebbene questa facoltà non sia un senso - egli scriveva - poiché non si relaziona con oggetti esterni, tuttavia è così simile alle facoltà sensoriali da potersi chiamare, con sufficiente proprietà, *senso interno*".

Nonostante l'indubbio interesse filosofico per tale "senso", il suo studio è stato storicamente snobbato dalla scienza, forse anche perché difficile da misurare.

La concezione illuministica dell'essere umano diffusasi in Europa a partire dal Settecento, per quanto riconoscesse l'importanza e la peculiarità del "senso interno" dell'autoconsapevolezza (già peraltro assurta a vero e proprio strumento filosofico grazie "cogito" cartesiano), non riteneva che esso si potesse studiare con metodi scientifici.

D'altra parte, in piena sintonia con tale concezione illuministica, ancora in pieno Ottocento, nel suo *Corso di filosofia positiva* (1830-1842) Auguste Comte scriveva: "L'individuo pensante non può tagliare se stesso in due - una delle due parti che ragiona, e invece l'altra parte che sta a guardare. Poiché in questo caso l'organo osservato e l'organo che osserva sono identici, come potremmo mai fare alcuna osservazione?"

Oggi, però le cose stanno profondamente cambiando: una nuova branca delle neuroscienze - la *neuroscienza metacognitiva* -, combinando innovativi test di laboratorio con le più recenti tecnologie di imaging cerebrale, pare stia finalmente sollevando il velo sulla mente che riflette su se stessa.

Dopo aver accertato che le tradizionali premesse comtiane erano fallaci, in quanto il cervello umano “non è” un singolo organo indivisibile, bensì una complessa rete composta da diverse strutture che contano miliardi di neuroni in continua interazione tra loro, Fleming ci propone un quadro scientifico particolarmente dettagliato di come funziona l’autoconsapevolezza in quanto processo cognitivo e neurobiologico, partendo dall’idea che nel cervello umano risiedono specifici “algoritmi di autoconsapevolezza”.

### ***Il calcolo dell’incertezza e l’automonitoraggio***

Tali algoritmi – sostiene lo studioso britannico – si formano a partire da due principi base. Il primo è connesso alla capacità del cervello di rilevare l’*incertezza* nella nostra percezione del mondo. “A causa della complessità del nostro ambiente e del fatto che i nostri sensi generano solo istantanee a bassa risoluzione dell’ambiente circostante, siamo costretti a fare delle ipotesi su cosa c’è realmente là fuori (...). Se fossimo incapaci di rappresentare l’incertezza, vedremmo (se mai) il mondo in un unico modo particolare; rappresentando l’incertezza, acquisiamo anche il nostro primo mattone di metacognizione, ossia la capacità di dubitare di ciò che i nostri sensi ci stanno dicendo”.

Una tale capacità di calcolare l’incertezza, tuttavia, è una caratteristica necessaria ma non sufficiente per raggiungere una piena autoconsapevolezza. A tal fine è imprescindibile la presenza di un’altra capacità, quella di *monitorare* le nostre stesse azioni.

Nel senso comune ogni azione umana è sempre considerata qualcosa di cosciente e deliberato, a partire dall’idea (in genere alquanto confusa) del *libero arbitrio*. Per capire quanto sia fallace un tale presupposto, è sufficiente riflettere sulla fondamentale importanza di tutta una gamma di possibili azioni che, ripetutamente, mettiamo in atto nel corso della nostra vita quotidiana: respirare, digerire, camminare, parlare, modificare la nostra postura. Tali azioni sono evidentemente inconsapevoli quanto essenziali per la nostra sopravvivenza.

Inoltre – e questo è un aspetto, secondo Fleming, fondamentale – anche le azioni più razionalmente e attentamente programmate, possono talvolta andare diversamente dalle previsioni. Sono perciò necessari sistemi che correggano i nostri errori con rapidità ed efficienza. Al cuore di questa nostra capacità di autocorreggere le nostre azioni vi è l’abilità di *predire* che cosa sarebbe potuto accadere, ma che non è accaduto. “La previsione – sottolinea Fleming – è in realtà al cuore degli algoritmi che possono monitorare se stessi” (possiamo soltanto riconoscere i nostri errori e rammaricarci dei nostri sbagli se sappiamo che cosa avremmo dovuto fare ma non abbiamo fatto).

La capacità di *monitorare* le nostre azioni può dunque essere considerato il secondo mattone necessario a edificare la base dell’autoconsapevolezza e dipende probabilmente da quello che Fleming definisce “un kit di partenza” neurobiologico presente precocemente nella nostra vita, condiviso peraltro anche con molte altre specie animali.

Se tali “mattoni” costituiscono le fondamenta su cui si erige l’edificio dell’autoconsapevolezza umana, altrettanto importanti sono quelle connessioni di mutuo rinforzo tra interazione sociale, linguaggio e ampliamento della capacità di modelli gerarchici ricorsivi e profondi, che progressivamente avrebbero condotto il cervello umano ad acquisire una capacità unica per la metacognizione cosciente, vale a dire la consapevolezza della nostra mente. “Questa forma di metacognizione si sviluppa lentamente nell’infanzia, e persino da adulti è sensibile alle variazioni di salute mentale, dei livelli di stress, e dell’ambiente sociale e culturale”.

### **Metacognizione e illusioni narrative**

A questo punto del suo lavoro la parte più originale del saggio di Fleming si occupa di provare a spiegare il ruolo dell'educazione e dell'apprendimento nei processi di elaborazione della metacognizione. Nel discutere tali questioni, l'aspetto teoricamente più significativo che emerge è quello relativo al rapporto tra *autoconsapevolezza e narrazione*.

Quando ci troviamo nella situazione di dover spiegare a qualcun altro i motivi che ci hanno spinto a scegliere un comportamento piuttosto che un altro, il più delle volte ricorriamo ad una *narrazione*. Citando celebri ricerche neuroscientifiche (come, ad esempio, quelle di Gazzaniga e Perry legate alla cosiddetta *sindrome del cervello diviso*), Fleming spiega in modo molto efficace come e perché le *narrazioni* siano così fondamentali per giustificare e dare un senso alle nostre azioni, indipendentemente dal fatto che le storie che narriamo (agli altri o anche a noi stessi) possano essere in parte, o addirittura del tutto, *fittizie*.

Se questo è vero (come, a quanto pare, confermano anche molteplici altre ricerche), la metacognizione ci appare in tutta la sua fragilità e inclinazione alle illusioni. In tal modo ci ritroviamo però di fronte a un insidioso paradosso: come fondare il nostro libero arbitrio?

*Come possiamo conciliare la tensione tra la metacognizione come supervisore imperfetto e come spiegazione del comportamento con il suo ruolo decisivo nel segnalare la responsabilità delle nostre azioni?*

L'autore fornisce due tipi di risposta per ridurre il peso di tale paradosso. La prima è che "non dobbiamo preoccuparci, perché il nostro senso di agentività sulle nostre azioni è accurato a sufficienza per buona parte degli scopi". *Nella maggior parte dei casi la metacognizione funziona in modo naturale, e abbiamo proprio una percezione accurata di cosa stiamo facendo e del perché la stiamo facendo* (dovremmo ammirare il fatto che il cervello possa fare tutto questo).

La seconda risposta è "riconoscere che l'autoconsapevolezza è soltanto un utile marcatore di responsabilità in una cultura che ammette che ciò sia vero. La nostra concezione di autonomia, come i nostri sistemi giuridici, è formata a partire dallo scambio sociale. La responsabilità, allora, è piuttosto affine al denaro. Questo ha valore soltanto perché noi siamo collettivamente d'accordo che l'abbia.

Analogamente, poiché siamo collettivamente d'accordo che l'autoconsapevolezza è marcatore utile di un particolare modo di prendere le decisioni, quest'ultima diventa centrale nella nostra concezione dell'autonomia. E, come il denaro, l'autonomia e la responsabilità sono in definitiva creazioni della mente umana, creazioni che dipendono dalla nostra capacità di concepire narrazioni su noi stessi e a vicenda. Come per il denaro - chiosa Fleming - possiamo riconoscere questa versione costruita della responsabilità, e al contempo godere del fatto di averla e di poterla usare".

E questo, indipendentemente dalla condivisione della consapevolezza del fatto che *l'introspezione sia, in definitiva, una finzione della mente*.

### **L'autoconsapevolezza nell'era delle macchine**

La fiducia in un'impostazione fondata su modelli di carattere quantitativo costituisce l'elemento più interessante (e, forse, più discutibile) dell'approccio di Fleming. Il suo presupposto è il seguente: al fine di creare una scienza dell'autoconsapevolezza, non è possibile affidarci "soltanto" a quello che ci dicono i

soggetti. Paradossalmente, infatti, se la nostra metacognizione è scarsa, è assai improbabile che noi lo sappiamo. È questo il motivo per cui – secondo questo studioso – si rende necessario il ricorso ad Intelligenze Artificiali coadiuvate da algoritmi elaborati a partire da ricerche di carattere quantitativo.

Già a partire dagli anni Sessanta, d'altra parte, si erano cominciati a sviluppare metodi per raccogliere e analizzare dati sistematici sui giudizi che le persone avevano di sé per poi compararli con la realtà delle loro prestazioni cognitive.

*L'idea si può riassumere più o meno così: dato che i circuiti cerebrali coinvolti nella creazione dell'autoconsapevolezza umana trascendono sia la percezione sia l'azione, combinando stime di incertezza tratte dai nostri sistemi sensoriali con informazioni relative al successo delle nostre azioni, è possibile intrecciare quelli che Fleming definisce i due mattoni fondamentali dell'autoconsapevolezza e analizzarli sulla base di "modelli" algoritmici.*

Oggi si sta lavorando a varie ingegnose soluzioni per creare *macchine introspettive*, che siano in grado di sapere se hanno probabilità di avere ragione prima di prendere una decisione, invece che a cose fatte. L'idea è quella di dotare queste intelligenze artificiali degli stessi mattoni metacognitivi che sono presenti in alcuni animali non umani e nei bambini, senza avere l'assurda (almeno a mio parere) pretesa di poterle dotare di un'autoconsapevolezza simile a quella umana.

Tutto sommato, se si evita di voler affrontare il cosiddetto problema "difficile" della coscienza (cfr. <https://www.exagere.it/riconoscersi-il-problema-irrisolto-della-coscienza>), i passi in avanti che consentono questo genere di ricerche ai fini di una sempre più raffinata conoscenza dei complessi meccanismi che caratterizzano la nostra autoconsapevolezza, risultano essere di straordinaria importanza.

Un'ipotetica macchina neurale per elaborare una conoscenza astratta su noi stessi – d'altra parte – dovrebbe essere presumibilmente molto simile a un genere di macchina utile alla creazione di una conoscenza astratta sul funzionamento di qualunque altro genere di cosa. Per riprendere un esempio introdotto dallo stesso Fleming, quando ci troviamo in un paese straniero e non sappiamo come utilizzare il sistema della metropolitana in cui non siamo mai entrati prima, ci aspettiamo, sulla base delle esperienze precedenti in metropolitane di altre città, che ritroveremo elementi simili (biglietterie automatiche, barriere e porte scorrevoli, indicazioni grafiche, etc.). Attingere a tali conoscenze pregresse accelererà notevolmente l'apprendimento del nuovo sistema. Ed è proprio un tale genere di astrazioni che, secondo gli studiosi della metacognizione, sarebbe necessario costruire all'interno delle reti neurali artificiali affinché le macchine "sappiano" che cosa "sanno" e che cosa "non sanno".

"Le convinzioni sulla mia personalità – scrive l'autore – le mie abilità e capacità mi aiutano a scoprire se sarò il tipo di persona capace di imparare una nuova lingua, di giocare a un gioco con la racchetta a me sconosciuto, o di fare facilmente amicizia. Questi fatti astratti relativi a noi stessi risiedono al vertice dei nostri modelli metacognitivi e, per via del loro ruolo nel plasmare come funziona il resto della nostra mente, esercitano una forza potente sulla nostra vita. È probabile che simili *autocredenze astratte* si riveleranno utili nel guidare robot autonomi verso compiti che si adattano alla loro nicchia...".

Sarebbe questa – sembra suggerire Fleming – la strada per renderli sempre più simili al personaggio dostoevskiano citato in apertura.

La conoscenza di noi stessi – dopo tutto – può essere considerata il tipo di conoscenza più astratta e

trasferibile in assoluto.

In fondo, pur consapevoli che si tratta di una mera *finzione grammaticale*, elaborata per motivi evolutivi dalla nostra mente, resterà sempre e soltanto il nostro specifico "io" l'unica caratteristica che potremo ritenere costante nelle diverse interazioni e situazioni con cui dovremo imparare ad adattarci nel corso della nostra esistenza.

---

Stephen M. Fleming

*Conoscere se stessi*

*La nuova scienza dell'autoconsapevolezza*

2022, Cortina Editore