

Idee per una rilettura
LA MENTE E IL COMPUTER,
INTRODUZIONE ALLA SCIENZA COGNITIVA

Il Mulino, Bologna 1990

di Philip N. Johnson-Laird

Gli esseri umani sono animali (Darwin)

Gli animali sono macchine (Cartesio)

Gli esseri umani sono macchine (La Mettrie)

Scopo del libro di Philip N. Johnson-Laird, psicologo inglese e professore di psicologia cognitiva e scienze cognitive presso l'Università di Princeton, è di "esaminare i processi mentali dal punto di vista della teoria della computazione" (pp. 6).

Il libro, pensato per studenti e per un pubblico interessato allo studio della mente dal punto di vista del paradigma *cognitivista*, è uno dei pochi manuali in circolazione che introduce espressamente ai metodi e ai modelli teorici delle scienze cognitive.

L'orizzonte teoretico di riferimento può essere individuato in quello che l'autore, in apertura, chiama un 'funzionalismo debole' secondo il quale si ammette la possibilità "che un programma per computer possa simulare aspetti dell'intelligenza umana".

L'autore, pur riconoscendo la natura non completamente computazionale della maggior parte dei processi mentali (visione, emozioni, udito, per esempio), resta un teorico dell'approccio simbolico in psicologia cognitiva, tanto da rievocare, in apertura, lo slogan caro ai funzionalisti e ai teorici di IA: "La mente può stare al cervello nello stesso modo in cui un programma può stare ad un computer" (pp. 28).

Naturalmente, ammette Johnson-Laird, si tratterebbe di una congettura destinata a rimanere tale, se i tentativi di risoluzione da parte degli scienziati cognitivi fossero di natura puramente *speculativa*. La questione, infatti, "è di natura empirica": costruire un robot in grado di percepire un mondo, interagire con esso ed esibire attività cognitive di alto livello.

Johnson-Laird non è solo uno scienziato cognitivo interessato alle questioni filosofiche ed epistemologiche della propria disciplina, ma è anche uno dei pochi psicologi sperimentali che scrive programmi per calcolatore con il preciso scopo di realizzare modelli computazionali delle sue teorie. Scrivere programmi al calcolatore per una data teoria psicologica, equivale, per lo scienziato cognitivo, a 'comprendere' i processi mentali alla

base del comportamento indagato (ragionamento, percezione, memoria, risoluzione di problemi).

L'idea che fa da sfondo a tutti i capitoli del libro è quella di *modello mentale*: una struttura cognitiva fondamentale per la comprensione dei processi percettivi (il risultato della percezione è un modello o sono modelli 3D del mondo) per la spiegazione dei processi di pensiero (il pensiero è la manipolazione *interna* di tali modelli che diventano, così, *mentali*) e per la formulazione di una teoria della competenza semantica (il linguaggio *serve* per comunicare il *contenuto* dei modelli mentali).

Nulla vieta, infatti, di dotare un robot dei più efficaci modelli mentali e delle procedure ricorsive in grado di manipolare tali strutture per ragionare, percepire e comprendere il linguaggio umano.

Se si dovesse scrivere, oggi, un libro introduttivo alle scienze cognitive, non potremmo fare a meno di dedicare almeno un capitolo alle neuroscienze cognitive, che in questi ultimi decenni hanno fatto veri e propri passi da gigante. Johnson-Laird fa sì riferimento ai contributi sperimentali offerti, per esempio, dalle teorie neuropsicologiche della visione, della coscienza, o allo studio dei famosi casi di 'split-brain' da parte del neuroscienziato Michael Gazzaniga, ma sempre privilegiando i modelli *computazionali* di tali attività cognitive.

Conosciamo le difficoltà che l'impostazione classica in scienza cognitiva affronta in sede di progettazione di automi o agenti artificiali già a un livello considerato basilare, come la percezione di oggetti, il riconoscimento di volti, o la comprensione del linguaggio. Consapevole di tali difficoltà, l'autore ammette che incorporare *in un* programma una teoria psicologica è, comunque, già un risultato significativo per lo scienziato cognitivo.

Oltre a illustrare i progressi della disciplina nei più importanti settori di indagine della mente (percezione, memoria, ragionamento, coscienza, linguaggio) il libro contribuisce a chiarire l'atteggiamento metodologico dello scienziato cognitivo nei confronti degli 'oggetti mentali'. Secondo Johnson-Laird, ciò che differenzia la scienza cognitiva dalle altre scienze, è il ricorso sistematico al computer, sia come metafora concettuale di riferimento nella comprensione dei processi mentali, sia come strumento in grado di incorporare, previa traduzione in un linguaggio di programmazione, una qualsiasi teoria della mente.

Una lettura più debole del paradigma funzionalista suggerisce, infatti, che un computer possa incorporare una teoria del mentale, o che sia in grado di simulare computazionalmente specifici processi mentali, senza, per questo, riconoscere che il

computer, o un ipotetico robot umanoide, siano dotati di un accesso interno ai propri pensieri (computazioni).

Un atteggiamento funzionalista ‘moderato’ non è detto, però, che sia quello corretto. Soltanto i progressi della robotica e dell’intelligenza artificiale potranno stabilire se un robot sarà in grado di possedere una vita mentale. I computer odierni non sono in grado di comprendere il linguaggio, perché non sono capaci di “metterlo in relazione con il loro dominio appropriato di interpretazione” (pp. 424). Ciò non esclude che in un futuro prossimo, risulti ancora impossibile “dare a queste macchine i meccanismi necessari per motivazioni, sensazioni interne e coscienza” (pp. 425).

La ‘computazione cognitiva’, aggiunge Johnson-Laird, suggerisce un’alternativa anche alle teorie dominanti in filosofia della mente: “I processi mentali sono le *computazioni* del cervello” (pp. 428).

Tale impostazione, sostiene l’autore, rigetta qualsiasi tipo di dualismo, assume l’esistenza di ben determinate caratteristiche della materia responsabili della possibilità di ‘esperire un mondo’, mette in primo piano l’organizzazione dei processi (mentali) e non esclude un’indagine sperimentale dei primitivi neurali (impulsi nervosi ed eventi elettrochimici in grado di produrre attività simbolica nella precisa struttura dei formati rappresentazionali). Ammette, inoltre, che non tutto è computabile in un senso classico e non rifiuta a priori che determinate computazioni, come quelle relative al pensiero deduttivo, possano essere conservate in un programma di un computer, così da costituire una ‘forma di rappresentazione dinamica dell’intelletto’ incorporabile in un modello mentale di un robot antropomorfo.

Tale impostazione sposa alcune congetture che non siamo obbligati ad accettare tanto facilmente: che la mente computi, o che il cervello computi, nel senso classico di Turing - von Neumann, non sembra affatto un risultato empirico *acquisito*. Che la maggior parte dei nostri processi cognitivi sia di natura computazionale non spiega come sia possibile concepire processi mentali non computabili: si pensi al settore della logica matematica e ai teoremi limitativi dimostrati dallo stesso Turing; se la mente di Turing è qualcosa che funziona come una macchina di Turing, allora, come spiegare il problema della fermata? Come hanno fatto le computazioni della mente *di* Turing a concepire un teorema che ‘dimostra’ la non computabilità dei suoi stessi processi mentali (delle sue stesse computazioni) se si accetta la computabilità di tali processi? Se un dato processo cognitivo è computabile fino a un certo punto, la parte non computabile è ancora da considerarsi un processo cognitivo nell’accezione che ne dà Johnson-Laird? E se non lo fosse, cosa

sarebbe? Di cosa scriviamo programmi? Di parti di processi mentali? Secondo questa impostazione, sarebbe più corretto definire la scienza cognitiva (classica) come la scienza che studia *le parti computabili* delle attività cognitive, ma se accettassimo tale 'condizionamento operativo' ne deriverebbe una visione molto circoscritta della mente.

In robotica, attualmente, viene data grande attenzione da parte dei progettisti ai contributi offerti dalle neuroscienze e dalla fisiologia animale. La ricerca attuale vede una sempre più marcata sovrapposizione degli studi classici sui sistemi artificiali con quelli basati sui sistemi naturali (si pensi alla 'cognizione incorporata' in IA, o ai robot realizzati dal ricercatore Rodney Brooks presso il Laboratorio di Intelligenza Artificiale del MIT). I sistemi artificiali odierni sono, però, molto lontani da fornire prestazioni accettabili in compiti 'apparentemente' privi di difficoltà per gli esseri umani come percepire, ricordare, o muoversi, il tutto intenzionalmente, ovviamente.

Le aspettative del cognitivista nei confronti della progettazione e della realizzazione di un robot che incorpori le più importanti funzioni cognitive umane rimangono, ad oggi, ancora inappagate. Per essere obbiettivi, l'autore del libro, è sì un cognitivista classico, ma è anche, potremmo aggiungere, un cognitivista 'critico'. I robot attuali, sottolinea a più riprese Johnson-Laird, non sono in grado di manifestare attività cognitive intenzionale, perché non sono capaci di mettere in relazione il livello delle computazioni con la varietà e la ricchezza sensoriale del mondo percepibile. Se, però, è possibile studiare la mente e le sue 'risorse' (termine recentemente introdotto da Marvin Minsky nel suo ultimo libro *The Emotional Machine*, per sostituire il ben più noto termine di 'agenti' comparso in *The Society of mind*) attraverso diverse classi di algoritmi, cercando, nel particolare comportamento di determinate computazioni la spiegazione dei fenomeni mentali, realizzare fisicamente queste computazioni in un robot diventa un banco di prova basilare per il teorico e per il progettista di agenti cognitivi artificiali (la mente sarebbe individuabile *nelle* computazioni del nostro agente artificiale).

Perché non denominare, un po' provocatoriamente, tale visione della mente, 'artificial behaviorism'? Resta da capire come un robot possa sviluppare autonomamente una teoria dei propri processi computazionali, ovvero, come possa *maturare* una psicologia se in lui esiste soltanto il livello delle computazioni. Tale argomento varrebbe anche se il robot fosse dotato di un sistema di interfacce visive, cinetiche, uditive, o altro; anche se possedesse delle istruzioni per costruire modelli mentali a partire da dati di tipo percettivo, o motorio, così da 'interpretare' le sue computazioni in un modello 3D del mondo, il robot non sarebbe in grado di sviluppare un accesso interno alle proprie

rappresentazioni: per fare questo, occorrerebbe inserire tali stati epistemici nella sua base di conoscenze in un formato simbolico o sub-simbolico, scegliendo un tipo di logica opportuna, o altre strutture come frames, script o appropriate reti neurali; ciò nonostante, i suoi modelli mentali continuerebbero a ignorare di possedere tali stati epistemici.

Dovremmo, inoltre, supporre che le computazioni cognitive siano, per il robot, 'sopravvenienti' rispetto agli impulsi elettrici che sfrecciano nei suoi microchip, ma questo non sembra essere sufficiente per sviluppare intenzionalità o memoria (a meno di considerare, per esempio, le schegge presenti in un sasso come 'memoria', nel senso di una proprietà fisica appartenente al sasso e ai *suoi* atomi). Questo è in parte corretto: ciò che chiamiamo memoria dipende da precise caratteristiche della materia e concordiamo con Johnson-Laird che questo valga anche per il pensiero; ma, se risulta *così fondamentale* il substrato che incorpora la 'natura dei processi psicologici', come facciamo a sostenere ancora la multi-realizzabilità del software mentale? Perché considerare soltanto i programmi per calcolatore, come l'equivalente dei processi cognitivi superiori negli esseri umani? Perché non considerare, per esempio, *pensiero* già i processi biochimici neurali? Non basta dire che l'intelligenza umana e artificiale si manifestano nella veste di un formato simbolico: un robot *non sa* di manipolare simboli, semplicemente esegue programmi installati dall'utente. I programmi di alto livello, così come le strutture rappresentazionali, non sono parte dell'ontologia interna del robot. Per il robot, non esiste neanche il livello delle computazioni, nel senso che *non sa* di stare effettuando computazioni, così come *per* la caffettiera non *esiste* il caffè come elemento intrinseco all'essere una caffettiera. Questo, si potrebbe argomentare, vale anche per noi umani: per esempio, la maggior parte di noi ignora il numero di cellule del corpo che stanno morendo in questo momento, ma possiamo dire di ignorarlo, appunto. Se il livello della manipolazione sintattica è ritenuto sufficiente per sviluppare un comportamento intelligente, non si vede perché non possano esserlo anche gli impulsi a livello delle sinapsi, o i fenomeni elettrochimici a livello molecolare presenti nei motoneuroni di una zampa di cavalletta, per esempio.

La mente è l'insieme dei processi sensoriali, percettivi, emotivi e cognitivi che il corpo è in grado di attivare. Quella tra mente e corpo è un'unità biologica vera e propria: la mente (umana e animale) semplicemente non esiste *priva* di un corpo; inoltre, il possesso di una mente non può inferirsi senza un comportamento osservabile e, si badi, il comportamento stesso non è soltanto qualcosa di osservabile, ma di *inferibile*.

Non sembra che la vita mentale si strutturi a partire da vuote occorrenze di un misterioso formato da riempire di volta in volta a seconda del dominio di esperienza considerato: il mondo che percepiamo non varia gran che, esso è la nostra forza di gravità, ciò che varia è il mondo interno (anche se non quanto vorremmo, purtroppo); ma se questa ricchezza di esperienza si colloca, per quanto ne sappiamo, a un livello 'subcognitivo' (forse citologico, o biochimico) cosa *intercettiamo* pensando? Dal punto di vista della vita, una scena percepita, un suono udito, una sensazione vissuta, un certo tipo di movimento, sono tutti ingredienti fondamentali per la genesi del pensiero, e, forse, sono già essi stessi *pensiero* puro (nel senso di *parti di mondo* sensoriale e percettivo *proiettate* nel corpo).

L'attualità del pensiero di Johnson-Laird, in linea con gli sviluppi di ricerca contemporanei, sta nel mettere in risalto, l'importanza della dimensione *interattiva* dell'intelligenza e della cognizione. Una cognizione, potremmo dire, condivisa, posseduta in comune con il mondo (in forma di modelli percettivi e mentali di questo) che fa da sfondo anche ai processi di ragionamento *astratto* e allo sviluppo della competenza semantica per i linguaggi naturali. Quando le aspirazioni del cognitivista si trasformeranno in realtà (robot antropomorfi) allora, i processi cognitivi *di tali* robot saranno molto simili ai nostri. Le operazioni mentali superiori coinvolte nella genesi di una teoria psicologica 'ingenua' saranno 'riproducibili' anche in un robot e un qualsiasi manuale di scienza cognitiva scritto per tale robot - che sarà in grado di aggiornare *automaticamente* i suoi modelli mentali collegandosi al web - non sarà, poi, così tanto diverso da quello offerto da Johnson-Laird.

Alberto Binazzi