

Roberto Cordeschi

Funzionalismo e modelli nella Scienza Cognitiva

(Sintesi di parte del cap. 7 di *The Discovery of the Artificial*, Kluwer Academic Publishers, Dordrecht, 2002)

In questo intervento mi propongo di discutere alcuni problemi epistemologici relativi all'uso dei modelli nello studio del comportamento degli organismi viventi, così come viene sostenuto dall'IA (Intelligenza Artificiale) interessata all'impresa multidisciplinare della Scienza Cognitiva. Mi riferirò all'IA e alla Scienza Cognitiva sia "classiche" che "nuove", ricorrendo solo per comodità alla distinzione che individua nelle prime l'approccio "simbolico" e nelle seconde gli sviluppi successivi all'approccio "subsimbolico" del connessionismo, quelli della Vita Artificiale, della *Animat Artificial Intelligence*, della robotica *behavior-based* ed evolutiva.

Vorrei mostrare come l'opposizione funzionalismo-antifunzionalismo, usata solitamente per contrapporre questi due approcci, è infondata: il funzionalismo è *parte* della strategia della costruzione dei modelli, e nella misura in cui quest'ultima è condivisa da entrambi gli approcci, non può essere visto come il peccato di origine della Scienza Cognitiva e all'IA classiche o simboliche. Devo precisare che mi riferirò al funzionalismo come alla tesi che i costituenti del modello (inorganico) possono avere lo stesso ruolo funzionale di quelli del fenomeno (organico) che si vuole spiegare, indipendentemente dal genere di realizzazione fisica. Non mi riferirò alla letteratura di filosofia della mente sul funzionalismo, quella che include autori come Putnam-prima-maniera, Block, Fodor, Searle e altri, e che è spesso riassunta nello slogan "la mente sta al cervello come il *software* sta allo *hardware*"—slogan completamente irrilevante, come dovrebbe essere chiaro nel seguito, per quanto riguarda i problemi epistemologici sollevati dalla costruzione dei modelli della Scienza Cognitiva, classica e nuova.

Detto molto in generale, la strategia della costruzione dei modelli nella Scienza Cognitiva parte da una teoria relativa a fenomeni che riguardano gli organismi viventi, come l'adattamento, l'apprendimento, la capacità di risolvere problemi, certe caratteristiche del loro sistema nervoso. La teoria si basa essenzialmente sull'osservazione del comportamento manifesto e, dove disponibili, su resoconti verbali o su dati biologici o neurofisiologici, e viene

usata come base per costruire un artefatto. L'artefatto, dunque, "incorpora", come spesso si dice, i principi esplicativi (o le ipotesi) della teoria, ed è considerato un criterio per il controllo della plausibilità della teoria.

Per chi accetta il metodo dei modelli nella Scienza Cognitiva si pongono, tra le tante altre, due questioni. La prima è quella della natura delle *restrizioni*, suggerite dalla teoria, che l'artefatto deve soddisfare perché esso sia un modello simulativo psicologicamente o neurologicamente plausibile, insomma degno di essere considerato per gli scopi della spiegazione del fenomeno indagato. Questo comporta che il semplice confronto del comportamento, o della prestazione, dell'artefatto con quello dell'organismo non basta per questi scopi. La seconda questione è la seguente: nel corso della procedura di controllo della teoria attraverso il modello, può essere necessario avviare un processo di *revisione*, che può riguardare tanto la teoria quanto il modello. In questo caso il modello assume una funzione euristica importante: esso può effettivamente suggerire nuove ipotesi da esplorare, e implicazioni rilevanti per la teoria stessa.

La storia della "scoperta" della metodologia simulativa che ha posto queste due questioni attraversa tutto il Novecento con il nome di "metodo sintetico". Essa parte da un'epoca anteriore di alcuni decenni alla nascita della cibernetica e arriva agli sviluppi più recenti, quelli dell'IA e della Scienza Cognitiva, classiche o nuove che siano. Tali sviluppi includono, ad esempio, i "modelli revisionati" di Allen Newell e Herbert Simon, il "processo dialettico" teoria-modello descritto da Philip Johnson-Laird, molti modelli della Vita Artificiale, quelli che incorporano la teoria del darwinismo neurale di Gerald Edelman, quelli della robotica *behavior-based*, evolutiva e in varia misura neurologicamente ispirata, certi sistemi multi-agenti e, infine, diversi cosiddetti "modelli sintetici completi" di organismi biologici molto semplici (Cordeschi, 2002). In tutti questi casi il modello è stato visto come un *terzo* modo di rappresentazione della teoria, che si affianca ai due più tradizionali, quello verbale e quello matematico. Attualmente, in alcuni casi questa rappresentazione della teoria ha la forma "virtuale" di un ordinario programma o di reti neurali e ambienti simulati al calcolatore, in altri casi la rappresentazione è data da un analogo fisico "realizzato", di norma un robot mobile. Altrove ho argomentato che questa distinzione non è importante in tutti i casi che riguardano la strategia dei modelli (Cordeschi, 2004). Il problema piuttosto è: quali requisiti deve soddisfare un modello, vuoi quando è una *simulazione* su calcolatore di un comportamento vuoi quando è un analogo *reale*, perché esso sia interessante nel contesto della spiegazione di tale fenomeno?

Questo interrogativo ci porta al problema del funzionalismo, la cui storia è parallela a quella della “scoperta” del metodo sintetico. Vediamo di eliminare alcuni equivoci prima di arrivare a chiarire qual è la relazione tra funzionalismo e metodo sintetico. A tale scopo prendiamo un esempio: la “modellistica neurale sintetica” di Edelman.

Edelman rifiuta il funzionalismo e accetta il metodo sintetico, che per lui include tanto la simulazione al calcolatore degli automi DARWIN quanto la realizzazione di robot mobili, come NOMAD. Edelman e collaboratori giustificano il rifiuto del funzionalismo sostenendo che l’approccio funzionalista al comportamento degli organismi lascia aperto il problema che i modelli da esso proposti “sono in linea di principio *sottodeterminati*” (Verschure et al., 1995, p. 247). Il problema della sottodeterminazione dei modelli è per loro il seguente: come varie macchine funzionalmente equivalenti o isomorfe non sono *distinguibili* per quanto riguarda la loro risposta, così, nella misura in cui il funzionalismo si limita a identificare una relazione tra lo stimolo e la risposta, tale relazione “risulta descrivibile da una molteplicità di modelli funzionali” basati su assunzioni diverse, i quali danno luogo a una stessa risposta e soddisfano le stesse predizioni, col risultato che “la scelta tra tali modelli alternativi non può essere fatta sulla base della sola risposta osservata” (p. 248). Per uscire dal vicolo cieco della sottodeterminazione in cui si caccia il funzionalismo, la proposta di Edelman e collaboratori è quella di costruire un modello (ormai sotto forma di un robot mobile come NOMAD) che tenga conto di restrizioni relative a componenti che influiscono sul comportamento manifesto dell’organismo a *più* livelli: almeno il fenotipo, il sistema nervoso, le proprietà dell’ambiente. Queste restrizioni, che risultano plausibili dal punto di vista biologico, dovrebbero garantire il modello dall’essere sottodeterminato.

Il problema posto dall’equivalenza funzionale ingresso-uscita, o stimolo-risposta, non è nuovo, e infatti Edelman e collaboratori esemplificano questa difficoltà del funzionalismo così inteso rifacendosi a un saggio pubblicato negli *Automata Studies* nel 1956 da Edward Moore sulle macchine isomorfe. Il funzionalismo della macchina di Turing risulta esposto, secondo Edelman, alla stessa obiezione della sottodeterminazione dei modelli, obiezione che coinvolgerebbe l’intera impresa della Scienza Cognitiva classica, fondata sull’asserito “paradigma della macchina di Turing”. Quello che però sembra sfuggire a Edelman, e che priva di forza la sua argomentazione contro il funzionalismo e a favore del metodo sintetico, è che nella Scienza Cognitiva classica il funzionalismo della macchina di Turing è stato ridimensionato da sempre, *proprio* per la difficoltà della sottodeterminazione dei modelli, dunque per le *stesse* ragioni esposte da Edelman e collaboratori. Il problema era in questo caso

quali e quante restrizioni imporre ai programmi perché potessero essere considerati modelli plausibili dei processi cognitivi, e non, come si esprimevano ad esempio Johnson-Laird o Zenon Pylyshyn, simulazioni *ad hoc* o meramente “mimiche”.

Quest’ultimo era particolarmente chiaro in proposito: “Se imponiamo a un processo l’unica restrizione di risolvere un problema in tempo finito, allora la tesi di Turing-Church stabilisce che, se il problema è solubile in modo meccanico, esiste una macchina di Turing (in effetti un’infinità di esse) che soddisfa i requisiti” (Pylyshyn, 1979, p. 43). Ma questa restrizione non è abbastanza forte da rendere, da sola, i programmi interessanti sotto il profilo della spiegazione di un fenomeno cognitivo. In realtà, quando nella Scienza Cognitiva classica è stato posto questo problema, non ci si è limitati a considerare solo la *prestazione* di un programma, come legittimamente si fa nel valutare, poniamo, un sistema esperto. Al contrario, una volta considerata la relazione funzionale tra lo stimolo e il comportamento in risposta osservato, ci si è posto il problema di identificare i *processi intermedi* che generano tale comportamento. A questo scopo è stato necessario considerare *altre* restrizioni, più forti, per i modelli cognitivi. Ai suoi tempi, Pylyshyn ne esaminava diverse: alcune erano state già individuate dalla ricerca precedente, mentre ad altre si era dedicata fino a quel momento poca attenzione.

Tra le prime Pylyshyn ricordava quelle popolari ai tempi della “simulazione del comportamento” proposta dalla *Information Processing Psychology* di Newell e Simon: dallo studio dei “protocolli ad alta voce” di soggetti umani alle prese con compiti di *problem solving*, Newell e Simon avevano ricavato diverse restrizioni che candidavano un programma ad essere una spiegazione. Si tratta di una fonte di restrizioni che nella sua forma originaria ha sollevato diversi problemi (Cordeschi, 1984), ma con essa si mostrava per la prima volta *esplicitamente* come il “livello di grana” della simulazione andasse specificato al di là della mera equivalenza funzionale. Tanto è vero che Newell e Simon respingevano come debole il test della mera equivalenza funzionale, nella interpretazione ristretta del test di Turing diventata poi popolare in IA, che ha dato luogo a varie incongruenze (la più nota e famigerata: l’effetto ELIZA). Tra le seconde restrizioni Pylyshyn indicava l’apprendimento, lo sviluppo, la stessa evidenza neurofisiologica, le proprietà dell’architettura cognitiva: tutte caratteristiche che, va notato, hanno costituito alcuni degli argomenti più controversi successivamente affrontati dalla Scienza Cognitiva fino ai nostri giorni. In quell’occasione, e in altre precedenti, Pylyshyn sosteneva che certe restrizioni provenienti dalle neuroscienze *potevano* risultare rilevanti dal punto di vista

della Scienza Cognitiva, come ad esempio i dati relativi allo studio di patologie come l'afasia e l'agnosia, o le ricerche sulla modularità di certe funzioni cerebrali (Pylyshyn, 1978, p. 126).

Successivamente, lo stesso Pylyshyn era tornato sul problema con l'obiettivo, però, di delimitare meglio il "genere naturale" della Scienza Cognitiva, e di conseguenza di precisare la natura delle restrizioni da imporre ai modelli computazionali. Ad esempio, egli escludeva che la coscienza o le emozioni potessero essere studiate come caratteristiche essenziali della cognizione (e si tratta di argomenti venuti invece in primo piano nella ricerca attuale). Egli poneva *proprio* la questione della sottodeterminazione dei modelli, nel senso già visto poi sollevato da Edelman, ritenendo che essa riguardasse i modelli cosiddetti "equivalenti in senso debole" alla cognizione umana, quelli basati appunto sulla mera equivalenza funzionale (ovvero sul comportamento osservato), solitamente richiamata facendo riferimento alla macchina di Turing. Si trattava di una prova di *sufficienza* dei meccanismi invocati per la realizzabilità della funzione, che però nulla diceva sui *processi cognitivi* implicati. Pylyshyn arrivava a rifiutare la stessa vecchia espressione "simulazione del comportamento" per qualificare i genuini modelli computazionali della Scienza Cognitiva. Questi dovevano essere "equivalenti in senso forte", vale a dire che per criteri di adeguatezza esplicativa occorreva individuare le restrizioni che indicassero la "corrispondenza a grana fine", come egli si esprimeva, tra processi cognitivi (umani) e processi computazionali (meccanici). Le restrizioni che egli individuava al livello dell'architettura cognitiva volevano essere una risposta a quanti ritengono che, dal momento che il criterio dell'equivalenza funzionale è debole, e i modelli restano in tal caso effettivamente sottodeterminati, le *uniche* restrizioni da imporre sono quelle biologiche, quelle che individuerebbero il *giusto* livello di grana per i modelli (Pylyshyn, 1984, p. 121). Si può dire che con quest'ultima mossa in quell'occasione egli toccava un tema, quello della *necessità* o meno delle restrizioni biologiche, che sarà poi portato in primo piano (si ricordi Edelman di sopra). Comunque, le sue conclusioni erano in quel momento scettiche circa l'utilità delle neuroscienze nella scoperta delle caratteristiche funzionali dell'architettura cognitiva.

Il ruolo centrale giocato dalle restrizioni è evidente in altri programmi di ricerca di Scienza Cognitiva, anche molto diversi tra di loro. Per esempio, per Johnson-Laird (1983) dal funzionalismo (da quello della macchina di Turing, evidentemente) deriva solo una "restrizione debole" sui modelli mentali, quella per la quale i modelli mentali sono computabili (p. 398); altre restrizioni riguardano, nel caso specifico e da lui meglio studiato del ragionamento sillogistico, criteri diversi: dal fatto che i modelli mentali spiegano errori o differenze individuali nella prestazione dei soggetti umani al fatto che spiegano come questi sono capaci

di fare inferenze valide, o come i bambini sviluppano tale capacità. Per fare un altro esempio, per Newell (1990) la proposta di un'architettura *unificata* della cognizione come SOAR significa proprio muoversi nella direzione di tener conto congiuntamente di una molteplicità di restrizioni. Queste restrizioni devono riflettere il fatto che la mente “si è evoluta in risposta a una molteplicità di esigenze diverse” (p. 18), tutte necessarie ad assicurarle un'interazione riuscita con l'ambiente, quali la flessibilità, l'apprendimento, le risposte in tempo reale, lo sviluppo, e così via, fino al fatto stesso di essere realizzata in un sistema nervoso. SOAR non le soddisfa certo tutte, ma la speranza è di riuscire a includerne nella sua architettura il maggior numero possibile. Quanto alle ultime restrizioni menzionate, le più vicine a quella che Newell chiamava la “*biological band*”, alcune gli sembrava fossero soddisfatte da SOAR (in primo luogo, quella della risposta in tempo reale); per il resto, egli concludeva che trascurare le restrizioni biologiche poteva essere giustificato ai tempi della *Information Processing Psychology*, ma non lo era ai giorni nostri, dati i progressi delle neuroscienze e la possibilità di costruire modelli che cominciassero a soddisfare quelle restrizioni. L'indifferenza verso le neuroscienze attribuita alla Scienza Cognitiva classica non costituisce dunque un tratto distintivo di ogni suo programma di ricerca.

Per riassumere. Il metodo dei modelli può essere considerato come il metodo che impone un certo numero di restrizioni agli artefatti *oltre* alla semplice equivalenza funzionale stimolo-risposta e alla semplice indipendenza della realizzazione in materia organica o artificiale. Nella sua generalità, questo metodo è stato condiviso tanto dall'approccio solitamente detto “dall'alto” attribuito alla Scienza Cognitiva e all'IA classiche quanto dall'approccio “dal basso” della Scienza Cognitiva e dell'IA nuove. Ogni ipotesi di realizzabilità in materia non organica di caratteristiche *essenziali* della mente, del cervello, del fenotipo, del genotipo degli organismi viventi condivide la strategia del funzionalismo. I modelli a reti neurali o della Vita Artificiale, i robot mobili, in quanto esempi di una metodologia modellistica o sintetica, condividono l'idea della rilevanza dell'organizzazione funzionale e della non rilevanza della materia organica nella spiegazione del mentale. Si legga ad esempio Langton a proposito della “tesi principe” della Vita Artificiale:

Il salto da compiere è di pensare alle macchine come a una questione di logica dell'organizzazione. *La materia non c'entra*. Non c'è niente di nascosto nella materia delle cose: una volta catturata la sua organizzazione logica in un altro medium qualsiasi, si ha la stessa “macchina”, perché è l'organizzazione che fa la macchina, non la materia di cui è composta (cit. da Levy, 1992, p. 117) [...] La tesi *principe* [della Vita Artificiale] è che un insieme opportunamente organizzato di primitive artificiali svolge lo stesso ruolo funzionale delle molecole biologiche nei sistemi viventi naturali. La Vita Artificiale è dunque *vita genuina*—semplicemente, consiste di materia diversa da quella che si è evoluta sulla Terra (Langton, 1989, p. 33).

Va ribadito che il funzionalismo non *basta* per gli scopi della spiegazione basata sui modelli (questo vale per il funzionalismo, ben precisabile, della macchina di Turing, e per quello, generico, della “macchina” della citazione di Langton). Esso stabilisce una relazione funzionale tra ingresso e uscita in sistemi dei quali si appura l’equivalenza della “logica dell’organizzazione”, per riprendere la vaga formulazione di Langton. Ma, e per mostrarlo ho voluto insistervi sopra, che esso non bastasse, che ci fosse (tra l’altro) il problema della sottodeterminazione dei modelli, lo sapevano anche i teorici classici del funzionalismo, i “funzionalisti della macchina di Turing” delle ricorrenti polemiche dei Churchland, Clancey, Edelman, Pfeifer, Searle e altri.

Mi sembra pertanto che ci sia un equivoco di fondo, largamente diffuso e riassumibile nello slogan seguente: l’IA modellistica *classica* è funzionalista (di qui il suo dualismo ecc.), quella *nuova* no. In realtà il funzionalismo non è un peccato originario più o meno da demonizzare. *Ogni* strategia modellistica è funzionalista nel senso sopra chiarito, e anche l’IA nuova non potrebbe non accettare l’idea della *comune organizzazione funzionale*: il tipo di struttura materiale non è direttamente rilevante nella comprensione del fenomeno (può trattarsi indifferentemente di composti organici o inorganici).

Quello che distingue i diversi indirizzi modellistici non è, allora, la scelta pro o contro il funzionalismo, ma è la scelta delle “primitive artificiali”, secondo la pittoresca definizione di Langton, ovvero del livello (psicologico, neurologico, biologico-evolutivo e così via) al quale il ricercatore decide di introdurre le restrizioni affinché il modello sia rilevante dal punto di vista della spiegazione. Per esempio, la Vita Artificiale tiene conto di restrizioni biologiche relative, di volta in volta, al genotipo, al fenotipo, all’ambiente, le quali *limitano* le simulazioni consentite. Ma anche per la Vita Artificiale vale il solito *caveat*, già evidenziato, come ho ricordato, dai cognitivisti classici a proposito dei loro modelli: “le simulazioni che dipendono da regole e restrizioni *ad hoc* o tagliate allo scopo, in quanto mimiche, non possono essere usate come sostegno di teorie della vita” (Pattee, 1989, p. 68).

Possiamo concludere che quella che la Vita Artificiale (o altre tendenze dell’IA nuova) rifiuta è la versione *computazionale* del funzionalismo, per la quale le restrizioni hanno interessato, soprattutto in passato (vedi i casi, molto diversi tra loro, della *Information Processing Psychology* o di Pylyshyn), prevalentemente, se non esclusivamente, il livello cognitivo alto. In realtà, chi condivide il metodo sintetico o metodo dei modelli, simulativi o

realizzati che siano, non può non condividere l'idea che i costituenti del modello (inorganico) hanno lo stesso *ruolo funzionale* di quelli del fenomeno (organico) che si vuole spiegare, indipendentemente dal genere di realizzazione fisica: di qui *si parte* per individuare il livello, o i livelli, che suggeriscono le restrizioni sui modelli.

Va ricordato che come non c'è unanimità su quali restrizioni dall'alto o cognitive imporre ai modelli, così non c'è unanimità su quali restrizioni dal basso o biologiche imporre ai modelli stessi. Sappiamo che i modelli connessionisti PDP o quelli basati su reti di Hopfield sono stati respinti da quanti sostengono i cosiddetti "modelli realistici del cervello", o dalla modellistica neurale sintetica di Edelman. Questi considera un numero sempre maggiore di aspetti anatomici e fisiologici del corpo (del fenotipo) e del sistema nervoso, e ora (con robot del tipo di NOMAD) anche dall'ambiente reale: tutte cose che impongono ai modelli restrizioni biologiche ben più esigenti di quelle che i connessionisti hanno proposto a partire dagli anni ottanta. In fondo, per Edelman i modelli computazionali a reti neurali non sarebbero meno *funzionalisti* di quelli della Scienza Cognitiva classica: i connessionisti PDP si illudono di proporre le giuste restrizioni biologiche, quando invece queste risultano ispirate a principi inaccettabili *proprio* dal punto di vista della plausibilità biologica, come per esempio la regola di apprendimento per retropropagazione e altre analoghe (Reeke et al., 1989).

In effetti, su questo punto Edelman ha ragione da vendere: proporre un'architettura a reti neurali o connessionista, dai tempi dei Perceptron e ai più progrediti modelli PDP o a reti di Hopfield, *non* equivale di per sé a soddisfare il requisito della plausibilità biologica. Possiamo cioè avere modelli basati sull'organizzazione funzionale dei neuroni in reti e i relativi meccanismi dell'apprendimento senza che tali modelli siano biologicamente plausibili, ovvero interessanti sul piano esplicativo. E la stessa questione della sottodeterminazione dei modelli, sollevata dalla mera equivalenza funzionale al livello simbolico (quello del programma), potrebbe essere posta dalla mera equivalenza funzionale al livello delle reti neurali, se non si invocassero le opportune restrizioni. (A meno di non voler riprendere vecchi slogan connessionisti, che oggi credo nessuno prenderebbe più sul serio, del tipo: "La semplice restrizione che i processi siano effettuati da neuroni interconnessi tramite sinapsi è incomparabilmente più forte di ogni altra restrizione accettata da altri approcci, in particolare dall'IA [classica]", Hanson e Olson, 1990, p. 2.)

La morale sembra essere che non esiste un criterio di decisione valido in generale per stabilire quali restrizioni prendere in considerazione. Piuttosto, queste sono suggerite dallo stato delle conoscenze sperimentali, dalle ipotesi rivali, dalla prestazione stessa del modello, dalla

scelta del livello di spiegazione da parte del ricercatore (Cordeschi, 2004). Una volta individuate, le restrizioni limitano quelli che si è disposti a considerare modelli plausibili ai fini della spiegazione, ma ogni modello, anche se è *realistico*, lo è sempre a un certo livello di astrazione o semplificazione della realtà. Come si esprimeva un pò paradossalmente Pylyshyn nel linguaggio dell'approccio dall'alto o cognitivo classico, “se impieghiamo il minimo delle restrizioni, avremo una macchina di Turing; se applichiamo tutte le restrizioni, può non esserci ragione alcuna di arrestarsi a non riprodurre un essere umano” (Pylyshyn, 1979, p. 49).

In conclusione, sostenere il metodo dei modelli nella Scienza Cognitiva vuol dire sempre avere come punto di partenza la tesi del funzionalismo. Quando ci si è posto il problema della plausibilità (psicologica o biologica) dei modelli, *tutti* i sostenitori del metodo dei modelli (non solo i cognitivisti classici) hanno dovuto fare i conti con i diversi problemi sollevati dal funzionalismo, da quello della sottodeterminazione dei modelli a quello della comune organizzazione funzionale a quello della realizzabilità in materiali sintetici, ai diversi e complessi problemi della spiegazione scientifica (si veda su quest'ultimo punto Tamburrini, 2001, e ancora Cordeschi, 2004). Come ho detto, dove i ricercatori si sono divisi è stato ed è nella scelta delle restrizioni da privilegiare per individuare, attraverso possibili soluzioni di quei problemi, il “genere naturale” dello studio della cognizione. In altri termini, essi si sono divisi nella scelta del livello o dei livelli di astrazione o di semplificazione ritenuti utili per la costruzione dei modelli. E' cioè utile il livello del simbolico, ritenuto proprio della Scienza Cognitiva classica, oppure quello subsimbolico delle reti neurali? E quanto è forte questa distinzione? Oppure dobbiamo considerare altri livelli, come quelli individuati dalla modellistica neurale sintetica di Edelman, o quello dei processi dello sviluppo scelto dalla Vita Artificiale o, su fronti ancora diversi, dalla robotica evolutiva, o quello dello studio dell'interazione tra agenti diversi? E in ogni caso, c'è una sostanziale differenza tra la *simulazione* al calcolatore di tutti questi processi, o di quelli dell'apprendimento e di altri ancora che si ritiene costituiscano caratteristiche essenziali del comportamento degli organismi viventi, e la loro *realizzazione* in agenti artificiali reali, secondo quanto affermano alcuni sostenitori della nuova robotica?

La risposta a queste domande ha a che fare con le procedure della ricerca scientifica, e non sembra possibile decidere quale sia la strategia migliore in *tutti* i casi che interessano la costruzione dei modelli del comportamento. Anche perché, al momento, le relazioni tra i livelli sopra menzionati non sono individuabili con facilità: in generale, si riesce a tener conto più o meno efficacemente di certe restrizioni biologiche quando ci si limita a modelli di prestazioni di

basso livello del vivente, mentre nel caso di prestazioni di alto livello è per lo più ineludibile il ricorso a restrizioni non biologiche. Inoltre, come ha detto efficacemente Harnad, se da un estremo i modelli basati sulla semplice equivalenza funzionale sono sottodeterminati, dall'altro estremo i modelli che si spingessero, poniamo, fino al livello della riproduzione sintetica dei neuroni nel maggior numero delle loro caratteristiche fisico-chimiche rischierebbero di essere "iperdeterminati", nel senso che, per essere il più possibile realistici, i modelli finirebbero per includere proprietà fisiche e funzionali che potrebbero essere irrilevanti o *non essenziali* (Harnad, 1994, p. 299) nello studio del fenomeno considerato: in questo caso i modelli finirebbero per oscurare quei principi o quelle ipotesi che sono chiamati a chiarire.

Le restrizioni aumentano effettivamente la possibilità che il modello possa dirci qualcosa sul fenomeno reale, ma non ci si può illudere che la regola sia sempre e solo: più sono le restrizioni, meglio è. A meno di non dover concludere, con la battuta di Norbert Wiener, che "il miglior modello del gatto è il gatto stesso".

Riferimenti

Cordeschi, R. (1984), "La teoria dell'elaborazione umana dell'informazione. Aspetti critici e problemi metodologici", in V. Somenzi (a cura di), *Evoluzione e modelli*, Editori Riuniti, Roma, pp. 319-420.

Cordeschi, R. (2002), *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Kluwer Academic Publishers, Dordrecht (tr. ingl. ampliata di *La scoperta dell'artificiale*, Masson/Zanichelli, Milano/Bologna, 1998).

Cordeschi, R. (2004), "Filosofia dell'Intelligenza Artificiale", versione on line prossimamente sul SWIF.

Hanson, S.J. e Olson, C.R. (1990), "Introduction: connectionism and neuroscience", in S.J. Hanson e C.R. Olson (a cura di), *Connectionist Modeling and Brain Function*, MIT Press, Cambridge, MA, pp. 1-4.

Harnad, S. (1994), "Levels of functional equivalence in reverse bioengineering", *Artificial Life*, 1, pp. 293-301.

Johnson-Laird, P.N. (1983), *Mental Models*, Cambridge University Press, Cambridge.

Langton, C.G. (1989), "Artificial Life", in C.G. Langton (a cura di), *Artificial Life*, Addison-Wesley, Reading, pp. 1-47.

Levy, S. (1992), *Artificial Life: The Quest for a New Creation*, London, Penguin.

Newell, A. (1990), *Unified Theories of Cognition*, Harvard University Press, Cambridge, MA.

Pattee, H.H. (1989), "Simulations, realizations, and theories of life", in C.G. Langton (a cura di), *Artificial Life*, Addison-Wesley, Reading, pp. 63-75.

Pylyshyn, Z.W. (1978), "Computational models and empirical constraints", *Behavioral and Brain Sciences*, 1: 93-127.

Pylyshyn, Z.W. (1979), "Complexity and the study of artificial and human intelligence", in M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence*, Harvester, Brighton.

Pylyshyn, Z.W. (1984), *Computation and Cognition: Toward a Foundation for Cognitive Science*, MIT Press, Cambridge, MA.

Reeke, G.N., Sporns, O. e Edelman, G.M. (1989), "Synthetic neural modelling: comparisons of population and connectionist approaches", in R. Pfeifer, Z. Schreter, F. Fogelman-Soulié e L. Steels (a cura di), *Connectionism in Perspective*, North Holland, Amsterdam, pp. 113-139.

Tamburrini, G. (2002), *I matematici e le macchine intelligenti. Unificazione e spiegazione nella scienza cognitiva*, Bruno Mondadori, Milano.

Verschure, P.F.M.J., Wray, J., Sporns, O., Tononi, T. e Edelman, G.M. (1995), "Multilevel analysis of classical conditioning in a behaving real world artifact", *Robotics and Autonomous Systems*, 16, pp. 247-265.