

Vecchi problemi filosofici per la nuova intelligenza artificiale

Roberto Cordeschi

cordeschi@caspur.it <http://www.scienze.com.unisa.it/hp%20cordeschi.htm>
Università di Salerno, Dipartimento di Scienze della comunicazione
I-84084 Fisciano (SA), Italy

Abstract. “Classical” AI supports the claim that representations have a central role in cognition. This paper deals with some philosophical problems regarding the role of representations in the modeling of natural agents in their interaction with the environment. The criticism of “new” or *nouvelle* AI to the representationalist claim of classical AI is debated through an examination of some situated robots or models, with the aim of evaluating the strength and the limits of that criticism. A different judgement on the role of symbolic representations in AI precedes some conclusions and suggestions for further research.

Sommario. Secondo l’IA “classica” le rappresentazioni hanno un ruolo centrale nella cognizione. In questo articolo si affrontano alcuni problemi filosofici relativi alla natura e al ruolo delle rappresentazioni nella modellizzazione di agenti naturali nella loro interazione con l’ambiente reale. Viene discussa la critica dell’IA “nuova” o *nouvelle* alla tesi sulle rappresentazioni dell’IA classica attraverso l’esame di alcuni casi di modelli robotici situati nei quali è possibile valutare la forza e i limiti di tale critica. L’esame di una diversa valutazione delle rappresentazioni simboliche in IA classica precede alcune conclusioni e qualche indicazione per la ricerca.

1 Introduzione

Che cosa sono le rappresentazioni? qual è il loro ruolo nelle teorie del comportamento degli esseri umani, degli animali e delle macchine? se ne può ipotizzare legittimamente o accertare l’esistenza a partire da dati relativi al comportamento di questi diversi agenti?

Si tratta di argomenti controversi, classici interrogativi filosofici che l’IA (Intelligenza Artificiale) si è posta fin dalle sue origini. Nella prospettiva di questo articolo, in tali interrogativi sono coinvolte, oltre a diverse tendenze dell’IA, da quella cosiddetta “classica” a quella cosiddetta “nuova” o *nouvelle* (sviluppi delle reti neurali nella Vita Artificiale, negli agenti autonomi, nella robotica situata, nella *Animat Artificial Intelligence*), anche altre discipline interessate all’impresa della scienza cognitiva, in particolare la psicologia, l’etologia e la neuroscienza cognitive.

Sia pure in via del tutto provvisoria, basandosi su Billman (1998), si possono introdurre preliminarmente le seguenti distinzioni a proposito dei molti modi con i quali si parla di rappresentazioni. Le *rappresentazioni esterne* sono, ad esempio, le mappe, che rispetto al territorio sono il sistema di rappresentazione. Le *rappresentazioni interne* possono essere ritenute presenti nella percezione, nel linguaggio, nel ragionamento, nella soluzione di problemi, e rappresentano nei sistemi cognitivi l’informazione proveniente dal mondo esterno, o elaborata da essi sulla base delle

proprie conoscenze, credenze, aspettative. Così intese, le rappresentazioni sono usate tanto da organismi, e si parla allora di *rappresentazioni mentali*, quanto da certi artefatti, come i programmi o i robot, e si parla allora di *rappresentazioni computazionali*. Le *rappresentazioni neurofisiologiche* sono viste come strutture neurofisiologiche, a diverso livello di complessità. Le *rappresentazioni teoriche o modelli* sono le rappresentazioni computazionali quando vengono impiegate come modelli della cognizione.

Nel corso di questo articolo mi riferirò implicitamente ad almeno alcune di queste distinzioni, e si vedrà come esse siano tutt'altro che scontate. Mi riferirò di volta in volta anche ad altri usi più particolari del termine "rappresentazione".¹ Questo articolo è così strutturato. Dopo aver riassunto la critica della IA *nouvelle* alle rappresentazioni nella versione dell'IA classica, prenderò in esame alcuni casi in cui si possono valutare la forza e i limiti di questa critica. La discussione di una diversa valutazione delle rappresentazioni simboliche in IA classica, ispirata a una versione di quest'ultima, precede alcune conclusioni e qualche indicazione per la ricerca.

2 Rappresentazioni e comportamento intelligente in ambienti reali

Il problema delle rappresentazioni, tanto nello studio del comportamento animale quanto nella progettazione dei nuovi robot cosiddetti autonomi, o *mobot* o *animat*, si è posto a proposito di come pattern spaziali possano essere codificati internamente da un agente, naturale o artificiale, e usati da questo per interagire efficacemente con un ambiente reale più o meno complesso.

L'interesse per compiti che si svolgono nell'ambiente reale è tipico dei sostenitori della nuova robotica, e sottolinea il loro distacco dall'IA classica, accusata di essersi troppo interessata alla simulazione di compiti artificiali (problemi giocattolo e micromondi), occupandosi dei problemi della "vita reale" solo nell'ambito dei sistemi esperti. In tutti i casi, l'approccio dell'IA classica a questi problemi viene rifiutato dalla nuova robotica in quanto *dall'alto*, dunque caratterizzato dall'interesse verso attività cognitive alte (pianificazione, soluzione di problemi nel gioco degli scacchi, nella dimostrazione di teoremi ecc.) e dal disinteresse più o meno pronunciato verso la componente percettiva e l'interazione senso-motoria dell'agente con l'ambiente. Anche quando l'IA si è occupata di quest'ultimo argomento, prosegue l'accusa, lo ha fatto ipotizzando, e realizzando con molta difficoltà e poco successo, agenti come quelli della robotica classica, che usano sistemi di rappresentazione della conoscenza esplicita e un controllo dell'azione che si svolge attraverso una pianificazione fortemente centralizzata.²

Queste critiche alla robotica classica hanno presente essenzialmente la robotica degli anni settanta, sempre esemplificata rifacendosi al robot mobile Shakey.

¹ Si parla, non di rado troppo genericamente, di rappresentazioni simboliche, esplicite, connessioniste, distribuite, implicite, quantitative, qualitative, metriche, topologiche. In ogni caso, nessuna di queste nozioni viene necessariamente usata in riferimento a un tipo di attività *soggettiva* cosciente (sul problema v. Trautteur 1995; Trautteur 2002).

² Anche David Marr, che è stato un critico determinato di questo tipo di IA, spostando l'attenzione della ricerca sulla percezione visiva verso i livelli bassi della visione, ha condiviso comunque l'approccio computazioni-su-rappresentazioni-simboliche. Qualche informazione su questi programmi di ricerca dell'IA (e su altri ancora ai quali mi riferirò in seguito) può essere trovata in (Cordeschi 1997) e (Cordeschi 2002, capp. 5 e 6).

Costruito negli Stati Uniti presso lo Stanford Research Institute, Shakey, nonostante il sofisticato sistema di pianificazione e rappresentazione della conoscenza, raggiungeva prestazioni modeste, confinate in un ambiente reale molto circoscritto e ben strutturato (nel quale, essenzialmente, era in grado di spostare alcuni grossi solidi). Le critiche alla robotica classica si riassumono nel rifiuto del cosiddetto “paradigma simbolico”, per il quale, dunque, la conoscenza del mondo (reale o simulato che sia) non è possibile se non attraverso la mediazione di rappresentazioni, e tali rappresentazioni, o simboli, sono strutture *linguistiche* manipolate attraverso processi di computazione che usano regole esplicite. Questo “paradigma” viene solitamente riferito all’Ipotesi del Sistema Fisico di Simboli, originariamente formulata da (Newell e Simon 1976), sulla quale tornerò nel seguito.

L’approccio *dal basso* proposto dalla nuova robotica vuole essere l’esatto opposto di quello simbolico-linguistico: si pongono al centro dell’attenzione i problemi dell’interazione senso-motoria dell’agente con l’ambiente, e si cominciano a studiare le sue semplici ma comunque efficaci strategie di adattamento e di apprendimento. La fonte prevalente di ispirazione è dunque non il comportamento di esseri umani nello svolgimento di compiti complessi, ma in primo luogo il comportamento di animali, anche semplici, ritenuti comunque “intelligenti” perché mostrano notevoli capacità di adattamento e di apprendimento in ambienti reali (spesso ostili) senza dover ricorrere a capacità cognitive alte, come quelle tipicamente studiate dall’IA classica. “Gli elefanti non giocano a scacchi”, è una efficace battuta di Rodney Brooks, inizialmente il principale ispiratore della nuova robotica. Uno slogan sempre di Brooks, “intelligenza senza rappresentazione”, fa riferimento a un’architettura dell’agente intelligente alternativa a quelle ritenute tipiche dell’IA classica. In tale tipo di architettura, dunque, il ruolo delle rappresentazioni non dovrebbe risultare cruciale.³

In breve, l’architettura proposta da Brooks, l’“architettura della sussunzione”, descrive l’agente come composto da più livelli di controllo funzionalmente distinti, che agiscono nell’interazione con l’ambiente in relativa indipendenza l’uno dall’altro, senza cioè la supervisione di un unico sistema centrale di controllo e di pianificazione dell’azione, basato su una rappresentazione o modello globale dell’ambiente o del mondo. I livelli superiori dell’architettura *sopravvengono*, per così dire, a quelli inferiori, subentrando nel controllo di un comportamento più complesso senza che vengano necessariamente interrotte le interazioni dell’agente con l’ambiente ai livelli inferiori. Ad esempio, l’agente può puntare alla ricerca di un obiettivo distante, o meta, e nello stesso tempo continuare ad evitare ostacoli diversi che incontra lungo la via. Questa architettura ha permesso di costruire robot mobili che (i) hanno doti di robustezza molto elevate rispetto a quelle dei sistemi di IA classica (sistemi esperti o robot) basati su rappresentazioni esplicite, essendo essi in grado di reagire efficacemente a situazioni non previste o non prevedibili, e (ii) sono capaci di interagire con l’ambiente in tempo reale, senza che sia cioè necessaria la complessa azione di pianificazione delle risposte tipica dei sistemi di IA classica.

Lo slogan originario di Brooks alludeva al fatto che tutta questa attività “intelligente” si svolge *senza rappresentazione*. In che senso?

In realtà, anche degli agenti di Brooks (delle sue “creature”, come egli le chiama) si può dire che hanno delle rappresentazioni. Allen, uno dei suoi primi robot, era

³ Accennerò in seguito ad altre posizioni antirappresentazionaliste e anticomputazionaliste, come quella di Tim van Gelder o quella dei costruzionisti, eredi delle vecchie critiche all’IA del filosofo heideggeriano Hubert Dreyfus o delle tesi autopoietiche di Huberto Maturana e Francisco Varela: si vedano rispettivamente van Gelder 1995) e (Winograd e Flores 1986).

munito di un sonar che gioca il ruolo di sensore di distanza: la ricezione dei segnali di ritorno funzionava come “rappresentazione interna” degli ostacoli che il robot doveva evitare, aggirandoli mentre esplorava un ambiente reale. L’espressione “rappresentazione interna” era usata dallo stesso Brooks (1991b: 231), ma altrove egli se ne mostrava insoddisfatto (“non mi piace chiamare rappresentazioni cose del genere”), perché il termine continuava a rimandare per lui all’intero armamentario dell’IA classica, come il modello centrale del mondo, l’elaborazione di piani, l’applicazione di regole esplicite, l’istanziamento di variabili e così via (Brooks 1991a: 149). Insomma, se quelle sono rappresentazioni, lo sono, si direbbe, in un senso molto *banale*, dal momento che la relazione dell’agente con l’ambiente e viceversa è così stretta e immediata che non è molto significativo dire che il primo si *rappresenta* il secondo.

La terminologia rappresentazionale, dunque, sembra risolversi per Brooks in un semplice modo di parlare, non caratterizzato da nessun intento esplicativo (non siamo tenuti, cioè, a introdurre le rappresentazioni come ipotetico meccanismo di *spiegazione* del comportamento dell’agente, una sorta di costrutti teorici intermedi tra stimolo e reazione): proprio come la terminologia mentalista che Brooks talvolta usa quando parla del “desiderio” di Allen di esplorare il mondo circostante, del suo “istinto” di evitare gli ostacoli e così via. Si tratta del resto dello stesso lessico usato in modo più o meno neutrale da Grey Walter (1953) nei confronti delle sue famose “tartarughe” o da Braitenberg (1984) nei confronti dei suoi ormai non meno famosi “veicoli”. Filosofi e psicologi hanno prontamente concluso che la nuova robotica originata da Brooks, spesso identificata come *behavior-based*, e in generale l’approccio dell’azione situata, hanno finalmente portato all’accantonamento della *disembodied mind* studiata dall’IA classica (funzionalista, dualista, cartesiana ecc.).⁴

Va precisato tuttavia che la ricostruzione sopra ricordata (approccio dall’alto contro approccio dal basso) poteva ancora essere proposta plausibilmente negli anni ottanta, ma attualmente, nonostante essa venga spesso riproposta in contesti diversi, non sembra più rispecchiare lo stato dell’arte, e rischia di dar luogo a una retorica dell’*embodied* non meno sterile del precedente asserito paradigma del *disembodied*⁵

In primo luogo, il panorama delle ricerche condotte al di fuori della robotica originata da Brooks si è molto evoluto e diversificato dai tempi dell’approccio dall’alto degli anni settanta. Nessuno (grazie anche alla lezione dello stesso Brooks) ha più in mente di costruire robot come Shakey, o comunque basati sulla netta distinzione tra componente percettiva e componente deliberativa e sul predominio della pianificazione centralizzata. Esistono ormai robot “ibridi” che manifestano buone caratteristiche di robustezza e azione in tempo reale, e nello stesso tempo usano diversi sistemi di pianificazione e rappresentazione della conoscenza classici (Carlucci Aiello et al. 2000). Un’illuminante rassegna di Bonasso e Dean (1997), che partiva

⁴ Sull’azione situata, le cui posizioni sono spesso molto vicine a quelle della nuova robotica, si veda, per gli scopi di questo articolo, (Clancey 1993), e come introduzione generale (Clark 1997).

⁵ Va detto che nessuno, anche nei primi tempi dell’IA, ha mai seriamente pensato che il “corpo” (la percezione, l’interazione con il mondo ecc.) non fosse importante per la cognizione. Piuttosto, si era stabilita una sorta di priorità, che portava a mettere in secondo piano questo aspetto, a vantaggio di modelli che soddisfacessero altre restrizioni: si vedano ad esempio le conclusioni in (Simon 1973). Qualcosa del genere vale anche per la restrizione del tempo reale: anche se si ammetteva che ancora non era stata affrontata, essa era considerata importante per i modelli della cognizione già nel classico articolo di Newell (1980) sui sistemi fisici di simboli.

dall'esame delle architetture robotiche alla Shakey, aveva già dato la sensazione dell'estrema varietà dei metodi sperimentati dalla ricerca robotica dopo Brooks, sia per gli aspetti relativi al controllo sia per quelli concernenti la natura delle componenti reattive, di quelle deliberative e della loro possibile integrazione reciproca, sia, infine, per quanto riguarda il ruolo svolto dal tipo di rappresentazione della conoscenza usato. Le stesse competizioni tra squadre di robot giocatori nelle RoboCup costituiscono uno straordinario laboratorio di sperimentazione e innovazione.

In secondo luogo, la robotica alla Brooks si è trovata in difficoltà nel dare sostanza ai ricordati slogan iniziali. Uno dei problemi che presto si pose ha riguardato il controllo in robot basati sull'architettura della sussunzione. Integrare efficacemente i diversi moduli componenti tale architettura a partire da quelli reattivi (in cui lo stimolo proveniente dall'ambiente genera l'azione, e questa dà luogo a sua volta a uno stimolo e così via) non risulta facile quando si tenta di rendere più elevata la prestazione di un robot. Proprio da questo tipo di difficoltà, che pone problemi di "complessità intrinsecamente esplosiva", è partita la robotica evolutiva (Cliff et al. 1993: 78). In effetti, uno sguardo all'evoluzione della ricerca lungo oltre un decennio mostra come la nuova robotica ha dovuto fare presto i conti con lo slogan radicale di Brooks ("intelligenza senza rappresentazioni"), e si è scontrata con una difficoltà che pure era stata prevista: quella di sviluppare le capacità dei robot da un livello meramente reattivo a un livello in cui sopravvenissero attività più complesse di interazione con l'ambiente, come qualche forma di pianificazione, di ragionamento, di apprendimento. Lo stesso Brooks è tornato successivamente su questa difficoltà, proponendo un nuovo slogan ("intelligenza senza ragione") contro l'asserito razionalismo ("cartesiano") dell'IA, ma con una precisazione di non poco conto per quanto riguarda le rappresentazioni, che, se confrontata con le sue iniziali affermazioni, sembra un vero ripensamento. La sua originaria critica alle rappresentazioni, egli sostiene, era diretta contro quelle introdotte dall'IA e dalla robotica classiche, non contro le rappresentazioni come "modelli parziali del mondo", cioè "relative al particolare compito per il quale [esse] sono usate", come egli si esprime alludendo alle prestazioni dei nuovi robot (Brooks 1995: 62).

Quello che mi interessa discutere in questo articolo è la natura di questi "modelli parziali del mondo", e come essi si collocano rispetto a quelle "rappresentazioni interne" che ho provvisoriamente chiamato *banali*, quelle che risultano dell'interazione senso-motoria diretta dell'agente con l'ambiente. Comincerò rifacendomi ad alcune ricerche di studiosi del comportamento animale dalle quali i sostenitori della nuova robotica hanno tratto ispirazione. Quelle più frequentemente menzionate sono le ricerche sull'ippocampo, tradizionalmente ritenuto deputato alla codifica neurale di rappresentazioni intese come mappe dell'ambiente (anche se attualmente il suo ruolo viene discusso da nuovi punti di vista), e in particolare le ricerche dei biologi (O'Keefe e Nadel 1978). Da queste si ricava che si possono assumere almeno due sistemi di navigazione nell'ambiente da parte di animali di diversa complessità: il primo riconducibile al "*taxon system*", l'altro al "*locale system*". Tra questi due sistemi i sostenitori della nuova robotica hanno collocato una serie di distinzioni (formulate nel lessico dell'IA classica e *nouvelle*: si veda ad esempio Prescott (1994; 1996) che, anche se non sono sempre univoche, proverò a inquadrare nella tassonomia seguente, che utilizzerò sempre in seguito per valutare diverse posizioni sull'argomento che mi interessa.

Il *taxon system* può essere visto come una catena di associazioni consistenti in coppie <stimolo, azione> che danno luogo al successivo stimolo, associazioni che consentono l'attività più semplice possibile di ricerca di cammini (*way-finding*). La

ricerca di cammini è una esplorazione-navigazione nell'ambiente che consiste nell'evitare o nell'avvicinare un determinato oggetto, seguendo segnali specifici (cioè rilevanti in una particolare situazione e per un particolare compito) provenienti dall'ambiente, oppure procedendo a contatto di una parete. Si tratta di un comportamento eminentemente reattivo, che dà luogo a una semplice forma di navigazione locale. Essa può comportare anche la capacità di memorizzare cammini, dando luogo a elementari forme di apprendimento specializzato (per rinforzo): si parla in questo caso di "apprendimento disposizionale".

Forme di ricerca di cammini più complesse della navigazione locale riguardano attività come la ricerca di un cammino verso una meta collocata fuori della scena visiva, o comunque non a portata dell'apparato sensoriale dell'animale. Questa ricerca di cammini, a sua volta, può essere di tipo rigido, limitato a prestazioni come la ricerca del nido o la ricerca di cammini familiari perché già attraversati in precedenza: una prestazione ben studiata, lo vedremo, già al livello di invertebrati come api o formiche. Oppure può comprendere rappresentazioni topologiche basate sulla presenza di punti di riferimento nell'ambiente (*landmark*) e relazioni topologiche come la maggiore o minore adiacenza spaziale tra tali punti di riferimento, l'ordine in cui occorrono e così via. Rispetto alla conoscenza dell'ambiente attraverso il *taxon system* (conoscenza data dalle associazioni di cui ho detto), la conoscenza basata su rappresentazioni topologiche ha un vantaggio: permette all'agente forme elementari di pianificazione di cammini, come quella consistente nel generare nuovi cammini usando segmenti di cammini diversi e già esplorati e nell'individuare così scorciatoie verso la meta.

Le rappresentazioni topologiche non permettono però l'individuazione di scorciatoie o cammini promettenti attraverso ambienti non noti o non esplorati in precedenza. Per questo ulteriore e più evoluto tipo di pianificazione di cammini sono necessarie informazioni su relazioni quantitative o metriche tra luoghi dello spazio, come la distanza relativa, quella definita rispetto a un quadro di riferimento e così via. Queste informazioni sono fornite dal *locale system* che è considerato un vero sistema di rappresentazione basato su una mappa che riproduce le relazioni metriche spaziali tra luoghi precisati dell'ambiente, la quale integra i dati disponibili in un modello unificato dell'ambiente. In questo caso la pianificazione di cammini può essere più accurata (con meno errori), più flessibile e selettiva, come avviene nei vertebrati superiori (negli invertebrati la cosa, vedremo, è discussa). "Apprendimento rappresentazionale" può essere detto quello che comporta rappresentazioni sia topologiche sia metriche (acquisizione di conoscenza causale e spaziale, presa di decisione attraverso inferenze ecc.).

In questa tassonomia si va dunque da un livello *egocentrico*, in cui l'agente è in interazione diretta con l'ambiente, a un livello *allocentrico*, dato da un quadro di riferimento più o meno globale. Per quanto riguarda i modelli della nuova robotica collocabili, per le loro prestazioni, ai vari livelli di questa tassonomia, va detto subito che, come osservano Wyeth e Browning (1998), mentre modelli che manifestano capacità al livello più o meno esclusivamente reattivo o di *taxon system* ormai abbondano, le prestazioni dei modelli che si muovono a un livello anche modestamente cognitivo variano enormemente in relazione al tipo di rappresentazioni che essi implementano, e al momento ciascun tipo pone problemi di non facile soluzione per il progettista.

Nel prossimo paragrafo esaminerò con qualche dettaglio tre casi diversi con lo scopo di far emergere alcuni dei principali problemi relativi alla natura delle rappresentazioni in relazione ai diversi livelli della tassonomia sopra descritta. Il

primo riguarda il caso di alcuni asseriti *robot naturali*, invertebrati come le formiche o le api, e costituisce un caso controverso: non è sicurissimo che in questo caso si possa invocare solo qualcosa di poco più che il *taxon system*, restando dunque a un livello esclusivamente egocentrico, per spiegare tutta l'evidenza comportamentale nota. Il secondo caso riguarda il comportamento di un robot *behavior-based* in grado di muoversi coerentemente in un ambiente reale, che sviluppa l'idea dell'architettura della sussunzione da un punto di vista prevalentemente egocentrico, ma mostra come sia necessario salire (sopravvenire) al livello topologico, e poi anche a quello metrico, per dotare l'agente di prestazioni appena più interessanti di quelle che vanno dalla navigazione locale (*taxon system*) a una pianificazione di cammini più o meno elementare. Il terzo caso propone, attraverso un robot solutore di labirinti, un modello soprattutto allocentrico, che fa riferimento esplicito alle mappe cognitive. Questi ultimi due casi si confrontano con il comportamento di un mammifero come il topo, traendo in diversa misura ispirazione dalle conoscenze biologiche e comportamentali note.

3 Agenti naturali e artificiali

Rudiger Wehner ha condotto con diversi collaboratori numerosissime osservazioni, lungo oltre un decennio, su un semplice invertebrato, una formica del deserto sahariano del genere *Cataglyphis*, creando situazioni sperimentali di sorprendente ingegnosità per l'interpretazione del suo comportamento. L'interesse per questo eccellente saggio di etologia cognitiva per quanto ci riguarda è che Wehner da una parte mette in dubbio la necessità di attribuire a sistemi organici semplici ma dal comportamento alquanto sofisticato come *Cataglyphis* le capacità computazionali necessarie per la costruzione e l'utilizzo di rappresentazioni sotto forma di una mappa cognitiva centrale, ma dall'altra non può fare a meno di invocare qualche forma di rappresentazione e di capacità computazionale per interpretare il loro comportamento.

Il comportamento della formica oggetto di studio è, in breve, il seguente. Una volta uscita dal nido alla ricerca del cibo, *Cataglyphis* se ne allontana, anche fino a duecento metri di distanza, esplorando il terreno attraverso un cammino molto irregolare e tortuoso, per poi tornare dritta al nido, non appena catturato il cibo, attraverso un cammino apparentemente mirato all'obiettivo. L'ipotesi elaborata da Wehner sulla base dell'evidenza comportamentale è che *Cataglyphis* riesce in questa prestazione perché è dotata di una vera e propria bussola solare, che usa per misurare continuamente gli angoli di rotazione e le distanze coperte, integrando le componenti angolari e lineari del movimento: in questo modo mantiene un vettore aggiornato a ogni passo che punta sempre in direzione del nido e che le consente di indirizzarsi verso di esso in qualunque momento. Questo è il suo principale metodo di navigazione nell'ambiente, consistente nella navigazione stimata (*dead-reckoning*).

Potremmo collocare questo sistema di navigazione locale di *Cataglyphis*, completamente egocentrico, al livello della ricerca di cammini elementare della tassonomia sopra menzionata. *Cataglyphis* mostra però anche una forma di apprendimento che, sempre sulla base della nostra tassonomia, potremmo chiamare disposizionale. Wehner (1997a: 10) ha osservato infatti che *Cataglyphis*, una volta tornata al nido, azzerando in questo modo nella sua "memoria di lavoro" il cammino percorso, conserva di fatto traccia di questa esperienza, come egli dice genericamente, "in qualche altra memoria di riferimento di ordine superiore", dal momento che si osserva come essa tenda a tornare più volte in luoghi noti dove aveva trovato del cibo.

Oltre alla navigazione stimata, Cataglyphis usa un secondo metodo di navigazione nell'ambiente che le consente di tornare al nido, e che consiste nell'uso di informazione proveniente da punti di riferimento nell'ambiente, sia intorno al nido e al luogo del cibo sia lungo il cammino tra questi due luoghi. Qui entriamo in un "terreno minato", come ammette Wehner, perché questa prestazione sembra suggerire la costruzione da parte di Cataglyphis di una rappresentazione sotto forma di un qualche tipo di mappa dell'ambiente prevalentemente allocentrica, topologica se non addirittura metrica. Wehner esclude l'uso da parte di Cataglyphis di mappe tanto topologiche che metriche. Piuttosto, egli conclude che la formica usa uno "schema" visivo bidimensionale semplificato che confronta con i punti di riferimento nell'ambiente per trovare il cammino, il cosiddetto "*snapshot matching*".

Ora, questo secondo metodo di ricerca di cammini usato da Cataglyphis comporta la costruzione di una mappa? Solo se prendiamo il termine nel senso più generale o più debole di una "mappa dei percorsi", sostiene Wehner, nella quale la formica elabora solo l'informazione basata su punti di riferimento relativa ai luoghi del nido e del cibo (Wehner 1997a: 9). Dalla nostra tassonomia sappiamo che casi di ricerca di cammini più evoluti di questo comprendono forme di rappresentazioni topologiche, come l'adiacenza o l'ordine in cui occorrono i punti di riferimento nell'ambiente, le quali permettono all'agente di generare nuovi cammini (pianificazione di cammini). Potremmo collocare in questo secondo tipo la mappa che per Wehner sta al livello immediatamente sopra quella dei percorsi, cioè la "mappa di vettori": in questo caso l'agente è capace di associare diversi vettori del nido con particolari punti di riferimento relativi a luoghi di cibo diversi, computando *nuovi* cammini attraverso la somma o la differenza di vettori. Wehner esclude che Cataglyphis sia un agente del genere (essa usa i punti di riferimento, attraverso lo *snapshot-matching*, solo per percorrere cammini noti), e a maggior ragione esclude che essa usi un terzo, e superiore, tipo di mappa: la vera e propria "mappa metrica" allocentrica, nella quale il nido, i diversi luoghi del cibo e i diversi cammini si collocano in un quadro di riferimento comune o globale.

Charles Gallistel ha contestato da sempre l'interpretazione di Wehner, valutando talvolta in modo diverso la stessa evidenza comportamentale degli invertebrati studiati da Wehner (dopotutto siamo in un "terreno minato"). Va osservato che Gallistel riconosce di poter parlare di rappresentazioni di diversa potenza, in relazione alla "varietà delle operazioni combinatorie che vengono impiegate per ricavare conclusioni valide sul sistema rappresentato", cioè il mondo o l'ambiente ([Gallistel 1990]: 3). Esiste così, dice Gallistel, una rappresentazione "poverissima", nella quale viene impiegata una sola operazione, ad esempio l'operatore di identità che userebbero gli invertebrati quando imparano a individuare e distinguere pattern bidimensionali. Ed esistono rappresentazioni più ricche, nelle quali vengono usati operatori che stabiliscono relazioni metriche tra i rappresentanti che denotano i diversi pattern nell'ambiente. È questo il caso delle api negli esperimenti di J.L. Gould, i quali mostrerebbero come, accanto al meccanismo dello *snapshot-matching* sostenuto da Wehner, questi invertebrati hanno la capacità di impiegare anche, in altre occasioni, vere e proprie mappe cognitive (metriche e allocentriche). Diversamente dallo *snapshot-matching*, le mappe cognitive permettono alle api un'attività più evoluta di ricerca di cammini, quella di pianificare cammini verso mete diverse dal nido. In questo caso esse calcolano la rotta che dovrebbero seguire a partire dalla stima della

posizione attuale nella loro mappa e della posizione dell'obiettivo (Gould 1990: 92-94).⁶

Dunque, mentre Gallistel sarebbe disposto a parlare di rappresentazioni, magari "minime", anche in assenza di mappe cognitive, Wehner è disposto a parlare di rappresentazioni solo nel caso delle mappe cognitive. In questo senso la sua mappa dei percorsi sarebbe probabilmente per Wehner un tipo di rappresentazione, sì, ma di rappresentazione *banale*, un pò nel senso di Brooks sopra ricordato. Eppure quello di *Cataglyphis* è molto più che non un comportamento meramente reattivo, almeno nel senso che in essa è presente un'attività computazionale che si basa su conoscenze dell'ambiente, locali quanto si voglia, codificate nel suo sistema nervoso. Dunque, come giustificare la presenza in questi animali di semplici forme di conoscenza? E' davvero *banale* il tipo di rappresentazioni "minime" da esse usato?

Ora nel caso della navigazione stimata, la formica usa, nelle parole dello stesso Wehner, una "rappresentazione interna del cielo" o schema bidimensionale molto semplificato, che, codificato nel suo sistema nervoso, senza essere una copia fedele del mondo esterno, si rivela comunque sufficiente allo scopo di orientarsi nell'ambiente. Anche il secondo meccanismo di navigazione usato da *Cataglyphis*, basato su uno schema visivo bidimensionale dei punti di riferimento dell'ambiente, è molto imperfetto ma generalmente efficace. I processi di computazione implicati in questi meccanismi di navigazione inducono in realtà lo stesso Wehner a parlare di un "modello semplificato del mondo" di *Cataglyphis*, basato su una modesta quantità di conoscenze, come egli dice, di tipo generico circa le caratteristiche geometriche della volta celeste, conoscenze "codificate nel sistema nervoso dell'animale" (Wehner 1997b: 36). Esse non comprendono, per esempio, quella che l'emisfero cambia nel tempo durante le escursioni fuori del nido della formica: sarebbe una conoscenza ridondante, dal momento che tali escursioni durano generalmente pochi minuti.

Si direbbe insomma che siamo qui in presenza di rappresentazioni "relative al particolare compito per il quale [esse] sono usate", cioè a "modelli parziali del mondo", per usare la più recente terminologia sopra ricordata di Brooks. I diversi compiti come quelli risolti da *Cataglyphis* attraverso rudimentali strategie computazionali di ricerca di cammini gli esseri umani li risolverebbero ricorrendo a strumenti diversi, sulla base della conoscenza dei calcoli necessari e facendo riferimento a un sistema tridimensionale. La prestazione di *Cataglyphis* suggerisce l'esistenza di un meccanismo neurale specializzato alla computazione del vettore (nel caso della navigazione stimata) piuttosto che di una rappresentazione del mondo di tipo generale. Si deve concludere che, per quanto innate, locali e specializzate, queste rappresentazioni non sono tanto *banali* da non essere oggetto di meccanismi di computazione, approssimativi quanto si voglia ma certo non poco complessi.

Riconosciuto questo, si potrebbe discutere *quali* sono gli algoritmi effettivamente implicati. Dickinson e Dyer (1996) hanno discusso questo punto valutando la capacità di modelli diversi, simbolici e connessionisti, nel soddisfare le restrizioni che devono essere loro imposte dall'evidenza ricavabile dal comportamento di api e formiche nell'interazione con l'ambiente. Rifacendosi a Gallistel, essi assumono che in un modello simbolico la rappresentazione interna dell'ambiente è resa da una funzione matematica esplicita: si tratta di computazioni su simboli codificati nel sistema nervoso; gli insetti avrebbero insomma, essi dicono, un'"equazione in testa" che, nel caso dell'orientamento degli invertebrati, descrive approssimativamente la posizione

⁶ Su questo punto si veda l'analisi di (Tamburrini 2002, cap. 1).

del sole con i parametri appropriati, secondo precedenti ricerche dello stesso Wehner (p. 195).

Ora spesso questo tipo di modelli riesce a inquadrare in modo soddisfacente l'evidenza sperimentale nota, ma da qui a dire che il modello descrive effettivamente quello che succede nel cervello dell'insetto ci passa. Come ha sottolineato lo stesso Wehner a proposito della sua spiegazione computazionale della navigazione stimata di *Cataglyphis*, "non si deve scambiare un algoritmo che riesce a descrivere la prestazione di un meccanismo neurale con il meccanismo stesso" (Wehner et al. 1996: 132). L'alternativa proposta da Dickinson e Dyer per evitare di assumere che i processi neurali implementino la computazione di funzioni matematiche esplicite è di ricorrere a modelli connessionisti, nei quali pattern dell'ambiente sono rappresentati sotto forma di pattern di attivazione distribuita in reti di unità interconnesse. Dickinson e Dyer rilevano vari limiti della prestazione di modelli a reti neurali semplici (eredi dei *Perceptron*), mostrando come altri modelli, più complessi, danno risultati più soddisfacenti, ad esempio lo stesso modello usato da Wehner per *Cataglyphis*. Da questo punto di vista, dovremmo concludere che le rappresentazioni e le computazioni rifiutate da Wehner sono quelle dei modelli simbolici, non di quelli connessionisti, che egli potrebbe tranquillamente accettare.⁷

Non entro ulteriormente nel merito di questo problema, limitandomi a osservare come ormai i robotici, interessati come sono alla *realizzazione* di agenti autonomi o situati, siano insoddisfatti di questo tipo di modelli a reti neurali non meno dei modelli simbolici. Più recentemente lo stesso Wehner ha collaborato alla costruzione di un robot mobile che realizza il principio della bussola solare ipotizzato in *Cataglyphis*. Egli e i progettisti del robot sottolineano il vantaggio di questa metodologia rispetto ai modelli a reti neurali tradizionali: "Di norma nei modelli a reti [simulate] della navigazione stimata si assume che l'informazione sull'orientamento, la distanza coperta e la velocità siano disponibili a partire dai sensori o possano essere facilmente calcolate. [...] Tutto quello che deve fare una rete neurale è manipolare quantità astratte. All'opposto, se una rete è implementata in un robot, diventa immediatamente evidente che l'informazione non è qualcosa di dato, ma deve essere estratta dall'interazione sistema-ambiente. Inoltre, non c'è un sistema ben definito di addestramento e controllo" (Lambrinos et al. 1997). Comunque, invece di fermarmi a considerare le prestazioni di questo robot vorrei passare a discutere le prestazioni dei due robot ai quali ho accennato, che risultano più istruttive per i miei scopi.

Il primo è il robot progettato da Maja Mataric. Con questo robot, detto Toto, (Mataric 1990; 1995) si proponeva di mostrare come, a partire dal livello meramente reattivo di un'architettura della sussunzione alla Brooks, si possono sviluppare in un agente rappresentazioni relativamente complesse anche se specializzate, sotto forma di una mappa topologica, che esso usa successivamente per interagire con l'ambiente in modo più efficiente. Mataric lo chiamava inizialmente anche "topo robot", perché si proponeva di realizzare un modello realistico dal punto di vista biologico del funzionamento dell'ippocampo del topo, riferendosi agli studi già menzionati di O'Keefe e Nadel. (Non mi fermo a valutare questo aspetto del modello.)

Toto è montato su una base circolare con tre ruote, che gli permette di muoversi lungo un cammino continuo in un ambiente reale chiuso, come una stanza arredata

⁷ Posto, naturalmente, che si sia disposti a parlare di rappresentazioni (magari "subsimboliche") nel caso dei pattern connessionisti. Per una discussione si veda (Bechtel e Abrahamsen 1991).

con oggetti diversi. Esso manifesta un elementare comportamento reattivo consistente nel seguire linee di confine (si muove seguendo le pareti e le sagomature degli oggetti) e nell'evitamento (evita altri robot o persone in movimento nella stanza). I punti di riferimento dell'ambiente sono rilevati dal robot, sulla base dei suoi stessi movimenti, mediante sonar e una bussola. In realtà il robot è in grado di riconoscere un certo numero di tipi di punti di riferimento. Uno di questi, ad esempio, comprende i corridoi. Un corridoio è per il robot una coppia: movimento lungo una linea (più o meno) retta e ricezione di rilevamenti (della parete) a brevi intervalli temporali.

Mentre si muove nell'ambiente seguendo le pareti e la sagomatura dei vari oggetti, Toto non si limita all'attività esplorativa ma costruisce gradualmente una rappresentazione interna dell'ambiente sotto forma di un grafo: i nodi del grafo corrispondono ai diversi oggetti dell'ambiente, rilevati di volta in volta quali punti di riferimento, e gli archi corrispondono alla relazione di adiacenza tra tali oggetti. Punti di riferimento fisicamente adiacenti nell'ambiente sono codificati topologicamente come nodi adiacenti nel grafo. Il carattere *distribuito* di questa rappresentazione è riconoscibile dal modo in cui sono memorizzati i punti di riferimento, il suo carattere *qualitativo* è dato dal fatto che l'accuratezza della rappresentazione è piuttosto bassa, basata com'è su relazioni di tipo topologico. Date comunque le limitate prestazioni richieste a Toto, è proprio questo che caratterizza la sua robustezza rispetto a un robot più tradizionale, che sfruttasse sensori molto precisi e un controllo accurato della sua posizione nell'ambiente.

La mappa così costruita è usata da Toto per memorizzare la struttura generale dell'ambiente in modo da poter raggiungere in successive esplorazioni luoghi dell'ambiente che possono essere stabiliti come mete dal progettista, o in modo da poter scegliere scorciatoie verso mete che il robot aveva raggiunto in precedenza attraverso cammini più lunghi e meno diretti. Questo tipo di ricerca di cammini è un'attività non semplicemente reattiva, ed è implementato nel robot usando una variante della attivazione diffusa dei nodi del grafo (date cioè la meta e l'attuale posizione del robot, l'attivazione si diffonde nel grafo dalla meta verso tutte le direzioni, permettendo un calcolo delle distanze tra i punti di riferimento noti che danno luogo a cammini alternativi). In questo modo il robot, a un dato punto di scelta, è in grado di prendere il cammino più breve verso la meta.

Toto non sfrutta dunque l'armamentario dell'IA classica, cioè un meccanismo di inferenza che genera un piano sulla base di un modello preconstituito del mondo. Quest'ultimo è interpretato da Mataric in un senso specifico, cioè come una mappa metrica in una struttura dati centralizzata: una mappa che, per avere successo, deve essere molto accurata. Piuttosto che massimizzare l'accuratezza, Toto fa i conti con l'inevitabile incertezza e anche gli errori che esso può accumulare mentre acquisisce la conoscenza dell'ambiente, e l'uso di informazione prevalentemente topologica si rivela in generale sufficiente a sopperire a questa sua debolezza. Ma come riconosce Mataric, tutto ciò, se da un lato semplifica la computazione, dall'altro limita la prestazione del modello a un tipo di pianificazione di cammini non molto evoluto: quello dell'individuazione delle scorciatoie attraverso cammini precedentemente *già* attraversati. Per migliorare la prestazione del robot nella pianificazione di cammini le sue capacità dovrebbero avere una "estensione" che desse luogo all'acquisizione della capacità di "fare inferenze geometriche", come ammette (Mataric 1995: 177), in definitiva di elaborare una vera e propria mappa metrica. La componente metrica, limitata nel modello descritto al calcolo della lunghezza dei cammini nella loro pianificazione, dovrebbe essere potenziata se il modello dovesse svolgere compiti di pianificazione di cammini più complessa: ad esempio, il compito di stabilire

attraverso una qualche forma di “ragionamento”, dice Mataric (Mataric 1990: 173), dei nuovi cammini quali scorciatoie che non siano cammini precedentemente attraversati. Questo, tra l’altro, accrescerebbe la plausibilità biologica del modello.

Come osservano anche Wyeth e Browning (1998), il topo robot di Mataric (almeno nella versione qui discussa) si troverebbe presto in difficoltà se la meta non fosse collegata in qualche modo a una parete, proprio perché la sua strategia di esplorazione principale è limitata al seguire linee di confine e sagomature. Esso non sarebbe un solutore di labirinti complessi efficiente come un vero topo, e neppure come il loro modello di solutore di labirinti, l’altro robot sul quale vorrei ora fermarmi brevemente. Va notato che mentre la ricerca di Mataric vorrebbe implementare un artefatto che risponda a una certa plausibilità biologica, quella di Wyeth e Browning è più sensibile ai problemi dell’efficienza della navigazione dei robot in quanto tali: ad esempio, essi sono interessati a un robot che apprenda l’esplorazione del labirinto nel tempo più veloce possibile.⁸ Tuttavia, proprio questo aspetto li porta a tener conto da una parte delle caratteristiche della navigazione effettiva degli animali viventi, che restano comunque un’ottima fonte di informazioni per il robotico (anche Wyeth e Browning si riferiscono agli studi sull’ippocampo di O’Keefe e Nadel), e dall’altra del tipo e della potenzialità delle rappresentazioni impiegate. Essi propongono un topo-robot in grado di risolvere labirinti di notevole complessità, che usa una mappa allocentrica, che cioè registra le relazioni tra gli oggetti del mondo, piuttosto che le relazioni tra l’agente e l’ambiente, così come era suggerito dai famosi esperimenti di Edward Tolman sulle mappe cognitive dei topi.

Il labirinto da essi impiegato ha la forma di 16×16 caselle, variamente configurabili in modo da ottenere corridoi diversi, anche a zig-zag, tali cioè che possano essere attraversati dal robot in diagonale, per avanzare così più velocemente che non seguendo la sagomatura del corridoio (la meta è al centro del labirinto). Il robot, un carrello con due ruote motrici come effettori, ha dei sensori (due laterali e uno frontale) che gli permettono di percepire la distanza che lo separa dalle pareti (laterali e frontali), e ha anche un sensore odometrico (nelle ruote) che registra la velocità e la distanza percorsa, permettendogli di conservare traccia della posizione raggiunta attraverso la navigazione stimata.

L’architettura del modello prevede tre livelli che, al solito, trovano posto (almeno i primi due) nella nostra tassonomia. Il modello manifesta infatti: (i) un comportamento di basso livello, il “livello degli schemi”, ovvero un livello di controllo puramente reattivo (alla Brooks, come riconoscono gli autori) che genera l’azione attraverso la semplice interazione con l’ambiente, senza accedere alla memoria (questo livello potrebbe corrispondere al *taxon system*); (ii) un comportamento al “livello cognitivo”, che usa la mappa del labirinto per pianificare l’azione sulla base degli schemi del livello inferiore (questo livello potrebbe corrispondere al *locale system*); (iii) un “livello motivazionale”, che genera le mete, e dispone di una serie di “valori” che il modello può usare per valutare le strategie più appropriate. I tre livelli funzionano in parallelo, cosicché il robot si muove e risolve il labirinto in uno stesso tempo. Detto molto in sintesi (e limitandoci ai primi due livelli), gli “schemi” percettivi (primitivi) del primo livello, nella misura in cui non hanno accesso alla memoria e non fanno riferimento alla conoscenza del mondo, non sono

⁸ Questi robot solutori di labirinti hanno i loro progenitori in era cibernetica e precibernetica (Cordeschi 2002, capp. 3 e 5). Il robot di Wyeth e Browning era stato progettato per partecipare a gare internazionali nelle quali diversi robot competono nel trovare il cammino più breve che li porta al centro di un labirinto. Ormai questo tipo di competizioni tra robot riscuotono meno interesse rispetto a quelle di Robocup.

rappresentazioni interne *nel senso di un modello del mondo*. E' al secondo livello che il robot costruisce una mappa, una vera rappresentazione del mondo, che gli consente di avere successo nel suo compito.

La rappresentazione del mondo è in questo caso facilitata dalla stessa struttura del labirinto: le 16×16 caselle possono essere rappresentate in modo univoco e regolare. Mentre costruisce la mappa al secondo livello sulla base delle informazioni fornite dal livello percettivo (ad esempio, la presenza o l'assenza delle pareti di una casella), il robot, mediante un algoritmo che calcola le distanze da ciascuna posizione di partenza alla meta, apprende infine il cammino più diretto verso la meta stessa (quello che gli permette di raggiungerla più velocemente). Wyeth e Browning analizzano la forza e i limiti del sistema di rappresentazione allocentrico del loro robot in relazione ad altri agenti più o meno egocentrici, concludendo che i sistemi di rappresentazione studiati dalla nuova robotica devono ancora chiarire come far emergere dal livello reattivo le capacità *inferenziali* dei robot, delle quali aveva parlato anche Mataric.

L'idea implicita nelle conclusioni di Wyeth e Browning, che cioè, nel caso di agenti artificiali, possano coesistere forme di rappresentazione dello spazio diverse, alla base di diverse strategie di esplorazione dell'ambiente, potrebbe essere riassunta nella "concezione dei modelli multipli" di Prescott, per la quale in un agente artificiale possono coesistere più descrizioni complementari dello spazio, che non dovrebbero necessariamente escludersi a vicenda. Proprio come avviene nel caso di *Cataglyphis*, che è in grado di elaborare due differenti strategie specializzate: essa segue il metodo di navigazione basato sui punti di riferimento nell'ambiente una volta tentato il primo, quello della navigazione stimata (e per la verità ne tenta come "estrema risorsa", dice Wehner, anche un terzo che finora non ho menzionato, quello dell'esplorazione sistematica o esaustiva dell'ambiente). "Benchè alcuni compiti possano richiedere la creazione momentanea di un modello unificato [come una mappa], in generale le rappresentazioni soggiacenti possono rimanere distinte, in modo che procedure alternative, magari contrastanti, non vadano perse" (Prescott 1996: 97).⁹

4 Simboli e rappresentazioni

In generale, nella nuova robotica c'è la tendenza a ricorrere al vocabolario delle rappresentazioni solo quando, per così dire, non se ne può fare a meno, ovvero, più precisamente, quando si lascia il livello puramente reattivo (*taxon system*), nel quale il lessico delle rappresentazioni sarebbe *banale*, per passare a quello topologico (è il caso di Mataric) e, a maggior ragione, a quello metrico (*locale system*) o delle mappe cognitive (il caso di Wyeth e Browning). Nolfi e Floreano (2000: cap. 5) mostrano come robot puramente reattivi sono capaci di risolvere alcuni compiti per i quali, nella ricerca su animali reali da Tolman a Gallistel, si erano invocate rappresentazioni complesse come le mappe cognitive. Ma essi mostrano anche come questi stessi robot reattivi, man mano che si riducono le restrizioni sull'ambiente, diventano sempre meno abili nell'affrontare quegli stessi compiti, che possono essere risolti solo da agenti dotati di stati interni ai quali essi riconoscono lo status di rappresentazioni. Parafrasando quello che David Marr e Tomaso Poggio dicevano del loro lavoro sulla visione, la massima sarebbe in questi casi quella di esaminare tutti i modi possibili di "spremere l'ultima goccia di informazione" dal livello reattivo prima di parlare

⁹ Si veda anche (Clark 1997).

dell'influenza di rappresentazioni, modelli del mondo o mappe sul comportamento intelligente¹⁰

Circa la natura delle rappresentazioni, una volta ammesse, le opinioni sono contrastanti, e riflettono la varietà dei punti di vista ormai usuale in IA, classica o *nouvelle* che sia. Per tornare ad alcuni dei casi esaminati, per Dickinson e Dyer si può parlare di rappresentazioni anche per i pattern connessionisti, a patto di distinguere le relative computazioni. Per Gallistel le rappresentazioni sono solo simboliche, quale che sia la loro complessità, e i pattern connessionisti, non essendo considerati da lui simbolici, non sono rappresentazioni.¹¹ Si parla dunque, come si è visto, di rappresentazioni che possono essere di diversa complessità e accuratezza, esplicite o implicite, metriche o topologiche, centralizzate o distribuite. E in generale si parla di rappresentazioni simboliche quando si è in presenza di costrutti dotati di proprietà ritenute analoghe a quelle *linguistiche*.¹²

Ho ricordato all'inizio come ricorrenti valutazioni polemiche da parte di alcune tendenze dell'IA *nouvelle* identificano nell'Ipotesi del Sistema Fisico di Simboli (PSSH) il "paradigma" linguistico per eccellenza dell'IA classica. Tuttavia, un confronto di qualche anno fa tra sostenitori e critici della PSSH mostra come questa interpretazione sia quanto meno opinabile. Sarebbe opportuno tenerne conto, per evitare di porre in un modo troppo sbrigativo l'identificazione tra *simbolico* e *linguistico* in IA classica e per affrontare senza pregiudizi i difficili problemi che stanno alla base della costruzione dei modelli della vita mentale, tra i quali quello della natura delle rappresentazioni. Mi riferisco all'interpretazione data da Herbert Simon e Alonso Vera in termini di sistemi di elaborazione simbolica dell'informazione (dunque in termini di sistemi fisici di simboli) di sistemi tradizionalmente non considerati tali, come quelli proposti dai teorici dell'azione situata e quelli alla Brooks.

Detto in breve, l'idea di simbolo che sta alla base della PSSH è che un simbolo è un pattern che denota, e la nozione di denotazione è quella che dà al simbolo la sua capacità rappresentazionale.¹³ I pattern possono denotare altri pattern, sia interni al sistema che esterni ad esso (nel mondo reale), e anche stimoli sensoriali e azioni motorie. Processi tanto biologici quanto inorganici possono essere simbolici in

¹⁰ L'espressione di Marr e Poggio è citata da Fodor (1983, p. 118).

¹¹ Si veda per una formulazione particolarmente esplicita (Gallistel 1999).

¹² Detto in breve, tali proprietà riguardano, tra l'altro, la produttività, ovvero la capacità di generare e capire un insieme illimitato di frasi, e la sistematicità, ovvero la capacità di capire ad esempio tanto *aRb* quanto *bRa*. Fodor ne ha fatto la base per la sua controversa ipotesi del "linguaggio del pensiero" Per una introduzione all'argomento, si veda (Di Francesco 2002).

¹³ Per pattern si intende, come sarà più chiaro nel seguito, una struttura fisica, biologica o inorganica, che può essere oggetto di processi computazionali—codifica, decodifica, registrazione, cancellazione, cambiamento, confronto—i quali occorrono in sistemi diversi, in un calcolatore e nel sistema nervoso, anche se in quest'ultimo caso non sappiamo nei dettagli come. Questa tesi provocò diverse reazioni (si vedano i volumi 17 e 18 di *Cognitive Science*). Si noti che nelle intenzioni di Simon e Vera la tesi non comporta che *ogni* pattern sia dotato di meccanismo denotazionale, cosa che evidentemente renderebbe banale questa definizione di simbolo: ci sono pattern che *non* denotano, tanto naturali quanto artificiali (si veda Vera e Simon 1994, p. 356). Sulla sufficienza della denotazione per caratterizzare la nozione di simbolo (come di rappresentazione) si è molto discusso. Si veda (Bechtel 1998) per una discussione particolarmente utile ai fini del punto di vista sostenuto nel presente articolo.

questo senso e, dal punto di vista sostenuto da Simon e Vera, i relativi sistemi sono *sempre* sistemi fisici di simboli, ma a diversi livelli di complessità.

Per esempio, nel caso più semplice che riguarda gli organismi, anche l'azione riflessa (subcorticale) è un processo simbolico: la codifica di un simbolo provocata da un ingresso sensoriale, poniamo la bruciatura di una mano, dà luogo alla codifica di un simbolo motorio, con la conseguente rapida effettuazione dell'azione, in questo caso il ritirare la mano. Più precisamente, l'idea è che "il sistema nervoso non trasmette certo la bruciatura, ma ne comunica l'occorrenza. Il simbolo che denota l'evento [la bruciatura] viene trasmesso al midollo spinale, che a sua volta trasmette un simbolo ai muscoli, i quali esercitano la contrazione che consente di ritirare la mano" (Vera e Simon 1994: 358). Nel caso degli artefatti, già il solito termostato è un sistema fisico di simboli, sebbene particolarmente semplice: il suo livello di tensione è un simbolo che *denota* uno stato del mondo esterno (p. 357).

Come ho ricordato, anche Brooks ha finito per riconoscere alle rappresentazioni un loro ruolo nel comportamento dei suoi robot, se non altro alle rappresentazioni "relative al particolare compito per il quale sono usate" (i "modelli parziali del mondo"), quali potrebbero essere, a diversi livelli di complessità, quelle usate da agenti naturali come *Cataglyphis* o da agenti artificiali come *Toto* o il solutore di labirinti sopra ricordato. Simon e Vera considererebbero senz'altro agenti del genere come sistemi fisici di simboli, dotati di un'attività rappresentazionale molto sofisticata, anche se specializzata a un compito particolare. Ma essi includono tra i sistemi fisici di simboli anche artefatti molto più semplici, come il ricordato termostato, e agenti robotici puramente reattivi o collocabili al livello del *taxon system* (che, seguendo Prescott, era stato definito come una catena di associazioni consistenti in coppie <stimolo, azione>). Secondo i due autori, i primi robot alla Brooks sono (un tipo relativamente semplice di) sistemi fisici di simboli: anche l'interazione senso-motoria *diretta* di un agente con l'ambiente nella misura in cui dà luogo a un comportamento coerente alle regolarità dell'ambiente, non può essere considerata se non come manipolazione simbolica.

Ho ricordato sopra il semplice comportamento reattivo di Allen, che tramite sonar evita ostacoli presenti in un ambiente reale. In questo caso, i suoi ingressi sensoriali danno luogo a un processo di codifica, e i costrutti in gioco (i simboli, secondo la definizione sopra ricordata) che risultano da tale interazione sensoriale, e poi motoria, dell'agente con l'ambiente sono rappresentazioni interne (degli ostacoli esterni da evitare) in un senso non banale: l'informazione sensoriale captata dal robot è convertita in simboli, i quali sono manipolati al fine di determinare gli *appropriati* simboli motori che evocano o modificano un certo comportamento. L'assenza di memoria in questo tipo di agente comporta che l'azione sia eseguita senza una rappresentazione esplicita del piano e dell'obiettivo che orienta l'azione stessa (senza pianificazione), ma non che non ci sia attività rappresentazionale (simbolica).

Qual è la natura di questi *simboli*, di queste rappresentazioni *simboliche*?

Nel caso degli artefatti più semplici si tratta di rappresentazioni analogiche che stabiliscono e mantengono la relazione funzionale del sistema con l'ambiente. Questo, si è visto, è già vero per il solito termostato. Nel caso di Allen o di *Toto* e di altri robot simili (come pure di certi sistemi connessionisti, o che includono sistemi connessionisti), tali rappresentazioni (analogiche) hanno carattere temporaneo (senza intervento di memoria) e distribuito (non sono sottoposte a controllo centralizzato). In questi casi, una rappresentazione certo imprecisa ma sufficientemente efficace è fornita da un sonar sotto forma di un pattern interno fisico (un pattern di nodi della rete, nel caso di un sistema connessionista): essa denota o rappresenta per il robot un

ostacolo o una certa curvatura di una parete o di un percorso. Una volta che tale pattern venga comunicato a uno sterzo, esso determina l'angolo della ruota sterzante del carrello del robot.

Per quanto diversa a seconda dei casi, è sempre presente un processo di codifica-elaborazione-decodifica non banale, che stabilisce una ben precisa relazione funzionale tra il sistema e l'ambiente, e spiega il comportamento coerente dell'agente nell'interazione con il mondo. Non parlare di rappresentazioni interne, e limitarsi a dire che un agente "intrattiene certe relazioni causali con il mondo, non spiega *come* tali relazioni vengano mantenute. E' del tutto ragionevole sostenere che un agente mantiene l'orientamento verso un oggetto tramite una relazione causale con esso e che tale relazione è un pattern di interazione, ma non ha senso pensare che tale pattern venga prodotto per magia, senza un corrispondente cambiamento di stato rappresentazionale dell'agente, ovvero che esso possa aver luogo senza una rappresentazione [interna] fosse pur minima" (Vera e Simon 1993a: 40).

Rappresentazioni più complesse, che sono alla base di un'attività non semplicemente percettiva (*diretta*), sono presenti in altri casi, quando entrano in gioco la memoria, l'apprendimento, il riconoscimento di oggetti e l'elaborazione di concetti, la formulazione esplicita di una mappa o di piani alternativi, sotto forma di rappresentazioni *off-line*, e ancora. In molte di queste attività "alte" intervengono rappresentazioni esplicite, linguistiche e metriche, ma se si riconosce che la cognizione richiede questo tipo di rappresentazioni, è difficile mettere in dubbio che tali attività non condividono con attività più "basse" come la percezione, sulle quali esse vengono elaborate, il meccanismo denotazionale, sia pure in una forma minimale. A meno di restringere arbitrariamente la nozione di rappresentazione e di simbolo, non c'è ragione di riserarla esclusivamente a pattern linguistici, o ai costrutti della semantica denotazionale (variabili da vincolare ecc.). Penso si possa sottoscrivere questa conclusione di Bechtel: "la nozione base [di rappresentazione] è effettivamente minimale, tale da rendere le rappresentazioni più o meno ubiquie. Esse sono presenti in ogni sistema organizzato che si è evoluto o è stato progettato in modo da coordinare il suo comportamento con le caratteristiche dell'ambiente. Ci sono dunque rappresentazioni nel regolatore di Watt, nei sistemi biochimici e nei sistemi cognitivi" (Bechtel 1998: 313).¹⁴

Simon e Vera distinguono il livello della modellizzazione simbolica da quello della realizzazione fisica (sia biologica che inorganica) di un agente. Nell'interazione con l'ambiente, un agente ha un'attività rappresentazionale che è data dalle caratteristiche specifiche del suo apparato fisico di codifica-elaborazione-decodifica di simboli. Si pensi ancora alla codifica, molto approssimativa ma generalmente efficace,

¹⁴ Il riferimento di Bechtel al regolatore di Watt è polemico nei confronti di van Gelder (1995), che ne faceva il prototipo della sua concezione non computazionale e non simbolica della cognizione. In realtà questo tipo di artefatti analogici (sistemi a feedback negativo e servomeccanismi) erano stati interpretati come sistemi rappresentazionali già all'epoca della cibernetica, in primo luogo da Kenneth Craik, che ne aveva fatto la base per una "teoria simbolica del pensiero", come egli la chiamava, per la quale "il sistema nervoso è visto come una macchina calcolatrice capace di costruire un modello o un parallelo della realtà" (per dettagli si veda Cordeschi 2002, pp. 137 sgg.). Non entriamo in questa sede sui diversi problemi relativi al contenuto delle rappresentazioni, al ruolo dell'utente degli artefatti e alla natura della spiegazione cognitiva. L'articolo di Bechtel contiene una disanima efficace di questi problemi, rispetto a posizioni diverse come quella sostenuta da Clancey (1993) contro la tesi di Vera e Simon. (Sempre in Cordeschi 2002, cap. 7 si possono trovare argomenti convergenti con la tesi sostenuta da Bechtel.)

attraverso sonar degli ostacoli da parte di un robot reattivo, e alla relativa decodifica che si conclude in un ben determinato movimento. La modellizzazione simbolica di questa capacità non appare in linea di principio diversa da quella “alta” sopra ricordata. L’idea è che tutti questi tipi o livelli di rappresentazioni, da quelli legati alla percezione a quelli più alti della “ricognizione” (per usare un termine di Neisser), possono essere opportunamente modellizzati attraverso regole di produzione, come livello di descrizione di un sistema fisico di simboli.

Un robot basato sull’architettura della sussunzione non fa eccezione. Ad esempio, il funzionamento di un modulo reattivo al livello più basso dell’architettura, che controlla la reazione di evitamento di ostacoli, potrebbe essere reso da un’unica regola di produzione del tipo “*se c’è un ostacolo (rilevato attraverso sonar e bussola) allora fermati*”.¹⁵ Questa possibilità sembra essere stata presa in considerazione dallo stesso Brooks, che però la respingeva in questi termini: “Un sistema di produzione standard in realtà è qualcosa di più [di un robot *behavior-based*], perché ha una base di regole dalla quale se ne seleziona una attraverso il confronto tra la precondizione di ogni regola e una certa base di dati. Le precondizioni possono contenere variabili che devono essere confrontate con costanti nella base di dati. I livelli [dell’architettura della sussunzione] funzionano in parallelo e non ci sono variabili né c’è bisogno di tale confronto. Piuttosto, vengono estratti aspetti del mondo, che evocano o modificano direttamente certi comportamenti a quel livello” (Brooks 1991: 155).

Tuttavia, se distinguiamo il livello della realizzazione fisica da quello della sua modellizzazione, quella che Brooks chiama l’estrazione degli “aspetti del mondo” *rilevanti* per l’azione è descritta in modo adeguato da un opportuno sistema di regole di produzione, e tramite tale sistema un certo comportamento di una sua creatura può essere evocato o modificato nell’interazione con l’ambiente. E questo modello (a regole di produzione) delle regolarità comportamentali di diversi livelli dell’architettura della sussunzione può essere implementata in un dispositivo che, grazie all’elevato grado di parallelismo, presenta doti di adattività, robustezza e risposta in tempo reale paragonabili a quelle di un dispositivo *behavior-based* (Burattini et al., in preparazione). In questo senso, le regole di descrizione danno una modellizzazione adeguata del comportamento di un agente situato.

Oltre alle risposte *automatiche*, che nel caso dell’azione riflessa o “innata” e di quella reattiva possono essere rese attraverso un’unica regola di produzione (qualcosa che corrisponda a una relazione comportamentista $S \rightarrow R$), esistono le azioni *automatizzate* a seguito dell’apprendimento, quando cioè le regolarità relative a un certo comportamento sono state memorizzate, o quelle che comportano una relazione “diretta” con il mondo tramite le *affordance* alla Gibson.¹⁶ Un esempio sono le

¹⁵ In breve, le regole di produzione hanno la forma “*se... allora*”, o CONDIZIONE → AZIONE. La memoria a lungo termine di un sistema fisico di simboli è costituita da tali regole: gli antecedenti CONDIZIONE permettono l’accesso ai dati in memoria, codificati dai conseguenti AZIONE.

¹⁶ Le *affordance*, nella terminologia di Gibson (1986) sono invarianti dell’ambiente che vengono “colte” (*picked up*) dall’agente “direttamente” nella sua interazione con l’ambiente stesso, e “direttamente” viene interpretato come: senza la *mediazione* di rappresentazioni e di computazioni su esse. Un esempio sono i movimenti dell’agente in un ambiente nel quale deve evitare oggetti o seguirne la sagomatura e così via: un po’ quello che fanno i robot reattivi di cui ho parlato. L’esempio del termostato è ricorrente in scienza cognitiva e in filosofia della mente dai tempi della cibernetica (si pensi a Donald MacKay, Daniel Dennett, Zenon Pylyshyn, Jerry Fodor ecc.). E’ evidente che definire sistemi fisici di simboli artefatti di questo tipo (e del tipo dei robot di Brooks, come

risposte immediate che fanno seguito a sollecitazioni improvvise o impreviste provenienti dall'ambiente

Ora i teorici dell'azione situata (e, come si è visto, i nuovi robotici) insistono sul fatto che questi casi di interazione diretta con l'ambiente si svolgono in tempo reale, senza cioè che sia possibile quella presa di decisione, diciamo così, meditata che richiede la manipolazione di rappresentazioni e la pianificazione dell'azione. Si pensi all'esempio di Winograd e Flores (1986) dell'automobilista che, guidando, affronta una curva a sinistra. In primo luogo, secondo i due autori, non è necessario che egli faccia continuamente riferimento a conoscenze codificate sotto forma di regole di produzione—non è necessario *riconoscere* una strada per accorgersi che è “percorribile” (la “percorribilità”, questa è la tesi, è *colta* nella relazione diretta agente-ambiente). In secondo luogo, la decisione è presa dall'agente, per così dire, senza pensarci (senza pensare di posizionare le mani, di contrarre i muscoli, di girare lo sterzo in modo che le ruote vadano a sinistra ecc.). Tutto ciò avviene automaticamente e immediatamente, dunque senza applicare qualcosa come una successione di regole di produzione “*se... allora...*”. In conclusione, la tesi è che non è possibile modellizzare questo aspetto della presa di decisione istantanea, o in tempo reale, attraverso un dispositivo che comporta codifica-elaborazione-decodifica di simboli, dunque computazioni, regole di produzione e così via.

L'obiettivo della critica di Winograd e Flores è la teoria della presa di decisione nello spazio del problema, con il quale ha a che fare l'agente a razionalità limitata di Simon. Ora, se prendiamo sul serio la teoria di Simon, va detto che alla base del carattere *limitato* della razionalità dell'agente sta la complessità dell'ambiente non meno dei limiti interni dell'agente stesso (limiti di memoria, di conoscenza della situazione ecc.). Nel prendere la decisione, quest'ultimo, secondo la teoria di Simon, in generale non è in grado di considerare, come spazio delle alternative pertinenti, lo spazio di *tutte* le possibilità, ma solo una parte più o meno piccola di esso, e questa selezione avviene sulla base delle sue conoscenze, aspettative ed esperienze precedenti. Ora una presa di decisione istantanea, non meno di una presa di decisione meditata, è condizionata da questi elementi, i quali, una volta che abbiano indotto, poniamo attraverso l'apprendimento, la formazione di schemi automatici di comportamento (di risposte motorie, nell'esempio di sopra), finiscono per determinare l'esclusione *immediata* di certe alternative possibili (come, nell'esempio della guida, innestare la marcia indietro) a vantaggio di altre (come scalare marcia, frenare ecc.), e tra queste altre quelle suggerite dalla conoscenza dell'ambiente stesso (fondo strada bagnato ecc.) e dalle aspettative pertinenti.¹⁷ Secondo le stesse parole di Simon “il solutore di problemi non percepisce mai *Dinge an sich*, ma solo stimoli esterni filtrati attraverso i propri preconcetti” (Simon 1973: 199). Di norma, dunque, l'informazione considerata dall'agente non è collocata in uno spazio bene ordinato di alternative, generato dalla formulazione del problema: tale informazione è generalmente incompleta, ma è pur sempre sostenuta dalla conoscenza della situazione da parte dell'agente.

La proposta è, dunque, che la modellizzazione a regole di produzione di un'azione del genere, e in generale di una *affordance*, è un simbolo che, via il sistema percettivo di codifica, raggiunge la memoria del sistema per soddisfare la CONDIZIONE di una regola di produzione esplicita. In questo modo, soddisfatta la CONDIZIONE, si attiva

vedremo) comporta rinunciare al requisito dell'universalità per tali sistemi (sul quale si veda Newell 1980).

¹⁷ Su questo tipo di comportamento, che può essere visto in termini di “percezione attesa”, si veda (Datteri 2002, cap. 4).

la regola, e la produzione (la decodifica) del simbolo di AZIONE avvia la risposta motoria.

Da questo punto di vista, le *affordances* sono rappresentazioni di pattern del mondo esterno, ma con una particolarità: quella di essere codificate in un modo particolarmente semplice. Nell'esempio di sopra, *una volta che si sia imparato a guidare*, la regola è qualcosa come: “*se la curva è a sinistra allora gira a sinistra*”. Questa regola rappresenta la situazione al livello funzionale più “alto”, secondo la terminologia dei due autori, nel quale la rappresentazione che entra in gioco è “minima”. Un termine del genere, a proposito delle rappresentazioni, lo abbiamo visto usato da Gallistel, ma per Simon e Vera il termine rimanda alla forma della regola indicata, che può essere rapidamente applicata: in questo caso, cioè, non c'è bisogno di evocare i livelli “bassi” o soggiacenti, quelli coinvolti con l'analisi dettagliata dello spazio del problema e con l'applicazione delle opportune strategie di soluzione, che comportano computazioni generalmente complesse, sotto forma di successioni di regole di produzione. Questi livelli intervengono nelle fasi dell'apprendimento (quando si impara come affrontare le curve), e possono essere evocati dall'agente quando la situazione si fa complicata (si pensi a una curva a raggio variabile, che rivela la complessità dell'interazione codifica percettiva-decodifica motoria). E tanto un apprendimento imperfetto quanto una carenza, per i più svariati motivi, dell'informazione percettiva rilevante possono anche ostacolare l'accesso ai livelli soggiacenti che potrebbero dare luogo alla risposta corretta (non tutti coloro che hanno *imparato* a guidare riescono ad affrontare tutte le curve con pieno successo in ogni situazione possibile).

Insomma, in questa interpretazione di Simon e Vera l'interazione in tempo reale dell'agente con l'ambiente è data non dal fatto di essere non simbolica e di non poter essere modellizzata mediante regole di produzione, ma dal fatto di non dover accedere, per dare la risposta corretta, alla complessità delle procedure di elaborazione simbolica dei livelli soggiacenti a quello “alto”. E' nell'attività cognitiva ai livelli soggiacenti, allorché si elaborano piani e strategie di soluzione di problemi, che viene evidenziata la consapevolezza dell'agente.

Simon e Vera ponevano infine un problema che riguarda i limiti degli approcci reattivi, sul quale mi sono già soffermato, e che mi sembra condivisibile: “E' tuttora dubbio se questo approccio [*behavior-based*] si possa estendere alla soluzione di problemi più complessi. Le rappresentazioni non centralizzate e le azioni non pianificate possono funzionare bene nel caso di creature insettoidi, ma possono risultare insufficienti per la soluzione di problemi più complessi. Certo, la formica di Simon non ha bisogno di una rappresentazione centralizzata e stabile del suo ambiente. Per tornare al nido zigzagando essa non usa una rappresentazione della collocazione di ciascun granello di sabbia in relazione alla meta. [...] Ma gli organismi superiori sembrano lavorare su una rappresentazione del mondo più robusta, [...] una rappresentazione più complessa di quella di una formica, più stabile e tale da poter essere manipolata per astrarre nuova informazione” (Vera e Simon 1993a: 34).¹⁸ La successiva evoluzione della robotica sembra confermare questa osservazione.

¹⁸ La formica di Simon è un esempio diventato famoso, e in tempi recenti ripreso proprio come esempio di “situato” dallo stesso Brooks, da Christopher Langton e molti altri. Con esso Simon voleva sottolineare, come ho ricordato sopra, come la difficoltà della soluzione di un compito è data tanto dai limiti interni dell'agente quanto dalla complessità dell'ambiente esterno.

5 Per concludere, e continuare

Riassumendo, elenco quelle che mi sembrano alcune conseguenze della definizione di simbolo e di rappresentazione simbolica nella versione di Simon e Vera che andrebbero tenute presente ai fini di una discussione senza pregiudizi sulla natura delle rappresentazioni.

(1) Possono esserci azioni che non coinvolgono pianificazione (nelle quali è dunque *prevalente* il feedback, come nelle *affordance* e, al limite, nell'azione riflessa), ma il processo di codifica-decodifica-elaborazione di simboli (dunque la computazione su rappresentazioni) è sempre presente, sia pure nelle forme elementari o "minime" ricordate .

(2) i simboli non sono solo "strutture linguistiche": l'assimilazione dei simboli a "simboli linguistici" ha portato all'equivoco che per l'Ipotesi del sistema fisico di simboli, e per l'IA classica in generale, "linguaggio e meccanismi linguistici soggiacciono a tutti gli aspetti della cognizione" (Vera e Simon 1993b: 126).

(3) I simboli non sono solo simboli nel senso della logica, ad esempio variabili da vincolare attraverso costanti di un dominio o mondo (Vera e Simon 1993a: 37). Anche le rappresentazioni analogiche sono simboli, e come tali sono elaborate da processi computazionali.

Dunque, la cognizione non coinvolge un linguaggio del pensiero, non è necessariamente una sorta di *proof theory* cognitiva nel senso di Fodor. Tuttavia, anzi proprio per questo, da Simon e Vera viene il suggerimento che non esistono aspetti *non simbolici* della cognizione, e che non occorre cercare *criteri* per le rappresentazioni al fine di distinguere (se non sul piano della capacità rappresentazionale) l'attività percettiva dalla costruzione di piani, e in generale dalla cognizione "alta". Inoltre, anche le reti neurali sono pattern con un loro meccanismo denotazionale simbolico. Si tratta, come si vede, di tesi che aprono diversi interrogativi interessanti. Per esempio, esse suggeriscono una valutazione perlomeno inusuale del connessionismo e delle rappresentazioni distribuite. Per Simon e Vera anche le rappresentazioni analogiche soddisfano la composizionalità: che dire allora della distinzione simboli-subsimboli di Smolensky? e come valutare, nell'ambito dei modelli cognitivi, la (perlomeno attuale) non equivalenza di sistemi simbolici (classici) e reti neurali quanto a capacità di manipolare rappresentazioni strutturate?

Non è questa la sede per entrare nel merito di questi interrogativi. Resta il fatto che la posizione antirappresentazionalista radicale, proposta a partire dalla metà degli anni ottanta, e riassumibile nello slogan "intelligenza senza rappresentazione", a distanza di un decennio appare quanto meno ridimensionata. Al di là dei successivi ripensamenti o precisazioni di Brooks, penso qui a certe sue asserzioni come la seguente, che ha suggestionato non poca filosofia popolare sull'argomento: "Nei momenti in cui mi sento più radicale sono convinto che questo [che ci sia intelligenza senza rappresentazioni] risulterà vero per tutte le forme del pensiero, fino all'intelligenza umana" (Brooks 1991b: 238). Questo ripensamento risulta evidente tanto nell'interpretazione del comportamento di agenti naturali con prestazioni di un minimo di complessità quanto nella progettazione di agenti artificiali del tipo dei robot situati. In quest'ultimo caso, come è spesso accaduto nella pur breve storia dell'IA, sembra che si sia trattato di dover smaltire l'ebbrezza provocata dai primi successi di un nuovo programma di ricerca, quello degli *animat* appunto, rispetto a un altro programma di ricerca in precedenza dominante e ormai considerato in una posizione di stallo, quello dell'IA classica. Smaltita l'ebbrezza, sarebbe il momento di abbandonare anche la propaganda in tema di "paradigmi" contrapposti (quello

simbolico-linguistico e quello subsimbolico-non linguistico, o come si voglia), un altro tema preferito della filosofia popolare sull'argomento.

Ne segue, in particolare, che le critiche di cognitivisti della prima ora (Johnson-Laird, Marr) alla plausibilità di una mente *ecologica* gibsoniana sono tutt'altro che accantonabili: per quanto l'interazione con il mondo da parte dell'agente possa essere apparentemente "diretta" e coinvolgere, ad esempio nella percezione visiva, livelli bassi, essa è in realtà sempre abbastanza complessa da richiedere qualche forma di computazione su rappresentazioni¹⁹. Kirsh evidenziò tra i primi questo punto a proposito dei robot di Brooks, le cui posizioni ritenne vicine a quelle di Gibson. Kirsh argomentò in modo convincente che queste creature, diciamo così, gibsoniane, che estraggono *direttamente* informazione dal mondo senza la mediazione di rappresentazioni, difficilmente potrebbero arrivare a prestazioni di una complessità superiore a quelle di organismi molto semplici: "solo in un senso retorico i robotici [così egli chiamava i nuovi robotici] possono sostenere di ripudiare le rappresentazioni" (Kirsh 1991: 180).

Suggerisco qui per inciso e rapidamente, passando dai robot ai bambini, che una lezione (antiretorica) interessante viene dagli psicologi dello sviluppo. Alison Gopnik ha argomentato che i bambini cominciano a sviluppare molto presto, verso il quarto anno, un modello rappresentazionale della mente che consente loro di intrattenere con il mondo interazioni più complesse di quelle consentite dal proprio precedente modello, quello che percepisce la mente in rapporto "diretto" con la realtà, alla Gibson. Così, nei bambini verso il quinto anno d'età, "le funzioni psicologiche sono quasi tutte [...] mediate da rappresentazioni. Desideri, percezioni, credenze, simulazioni, immagini, condividono tutte la stessa struttura di base, quella che a volte viene descritta in termini di atteggiamenti proposizionali e di contenuto proposizionale. *Tutti questi stati mentali comportano rappresentazioni della realtà, piuttosto che relazioni dirette con la realtà stessa*" (Gopnik 1993: 6, corsivo mio).

Dunque, e tornando dai bambini ai robot, le rappresentazioni, dopo una breve euforia eliminativista, sono rientrate in gioco nelle forme più diverse, e la ricerca nel comportamento animale e umano, in IA (classica o *nouvelle* che sia) e in filosofia della mente, non potrà non tenerne conto. Per dirla ancora con Vera e Simon (1993a: 15-16): "Un compito complesso è molto più difficile da svolgere basandosi solo su un feedback diretto con l'ambiente piuttosto che, in più, su un modo per stabilire un qualche tipo di rappresentazione dell'ambiente stesso. Naturalmente, la pura pianificazione senza feedback di questo tipo si rivela ugualmente inefficace, ma è una sfortuna che i fallimenti della pura pianificazione abbiano indotto i ricercatori a passare all'estremo opposto, trascurando una strategia intermedia più sofisticata".

Ringraziamenti

Un versione di questo lavoro è stata pubblicata nel volume curato da U. Bottazzini e E. Di Bella, *Le costruzioni della mente*, McGraw-Hill Italia, Milano, 2001. Ringrazio Ernesto Burattini, Guglielmo Tamburrini e Giuseppe Trautteur, con i quali ho avuto la possibilità di discutere in tante occasioni gli argomenti qui affrontati. Anche con Giuseppe Boccignone, Luigia Carlucci Aiello, Edoardo Datteri, Marcello Frixione e Stefano Nolfi ho potuto discutere precedenti versioni di questo lavoro. Ringrazio anche loro, liberando tutti da qualsiasi responsabilità per le tesi in esso sostenute.

¹⁹ Questo può essere sostenuto anche senza aderire all'interpretazione delle *affordance* data da Simon e Vera. Si veda ancora Clark (1997).

Riferimenti bibliografici

- Bechtel, W. (1998), "Representation and cognitive explanations: assessing the dynamicist's challenge in cognitive science", *Cognitive Science*, 22, pp. 279-318.
- Bechtel, W. e Abrahamsen, A. (1991), *Connectionism and the Mind*, Oxford, Blackwell.
- Bechtel, W. e Graham, G. (eds.), (1998), *A Companion to Cognitive Science*, Oxford, Blackwell.
- Billman, D. (1998) "Representations", in Bechtel e Graham (1998), pp. 649-659.
- Bonasso, P. e Dean, T. (1997), "A retrospective of the AAAI robot competitions", in *AI Magazine*, 17, pp. 11-23.
- Braitenberg, V. (1984), *I veicoli pensanti*, Milano, Garzanti.
- Brooks, R.A. (1991a), "Intelligence without representation", in *Artificial Intelligence*, 47, pp. 139-159.
- Brooks, R.A. (1995), "Intelligence without reason", in Steels e Brooks (1995), pp. 23-81.
- Brooks, R.A. (1991b), "How to build complete creatures rather than isolated cognitive simulators", in VanLehn (1991), pp. 225-239.
- Burattini, E., Cordeschi, R. e Tamburrini, G. (in preparazione), "Subsumption architectures and symbolic processing for real-time action".
- Cantoni, V., Di Gesù, V., Setti, A. e Tegolo, D. (eds.), (2000), *Human and Machine Perception 3: Thinking, Deciding, and Acting*, Dordrecht, Kluwer Academic Publishers.
- Carlucci Aiello, L., Nardi, D. e Pirri, F. (2000), "Case studies in cognitive robotics", in Cantoni et al. (2000).
- Clancey, W. J. (1993), Situated action, in *Cognitive Science*, 17, pp. 87-116.
- Clark, A. (1997), *Being There*, Cambridge, MA, MIT Press.
- Cordeschi, R. (1996), "L'Intelligenza Artificiale", in Geymonat (1996), pp. 145-200.
- Cordeschi, R. (2002), *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Dordrecht, Kluwer Academic Publishers (ed. inglese ampliata di *La scoperta dell'artificiale*, Milano, Dunod-France/Bologna, Zanichelli, 1998).
- Datteri, E. (2002), *La percezione attesa: anticipazione e modelli dell'azione in tempo reale*, Tesi di laurea, Dipartimento di Filosofia, Università di Pisa.
- Di Francesco, M. (2002), *Introduzione alla filosofia della mente*, Roma, Carocci.
- Dickinson, J. e Dyer, F. (1996), "How insects learn about the sun's course: alternative modeling approaches", in *From Animals to Animals 4*, Cambridge, MA, MIT Press, pp. 193-203.
- Eliasmith, C. (1997), "Computation and dynamical models of mind", in *Minds and Machines*, 7, pp. 531-541.
- Fodor, J. (1983), *The Modularity of Mind*, Cambridge, MA, MIT Press (trad. it. *La mente modulare*, Bologna, Il Mulino, 1988).
- Gallistel, C.R. (1990), "Representations in animal cognition", in *Cognition*, 37, pp. 1-22.
- Gallistel, C.R. (1998), "Symbolic processes in the brain: the case of insect navigation", in Scarborough e Sternberg 1998, pp. 1-51.
- Geymonat L. (ed.), (1996), *Storia del pensiero filosofico e scientifico. Il Novecento*, vol. 8, t. 3 (a cura di E. Bellone e C. Mangione), Milano, Garzanti.
- Gibson, J.J. (1986), *The Ecological Approach to Visual Perception*, London, Erlbaum (trad. it. *Un approccio ecologico alla percezione visiva*, Bologna, Il Mulino, 1999).
- Gopnik, A. (1993), "How we know our mind", in *Behavioral and Brain Sciences*, 16, pp. 1-14.
- Gould, J.L. (1990), "Honey bee cognition", in *Cognition*, 37, pp. 83-103.
- Kirsh, D. (1991), "Today the earwig, tomorrow man?", in *Artificial Intelligence*, 47, pp. 161-184.
- Krebs, J.R. e Davies, N.B. (eds.), (1997), *Behavioural Ecology: An Evolutionary Approach*, Oxford, Blackwell, 4a ed.
- Lambrinos, D., Maris, M., Kobayashi, H., Labhart, T., Pfeifer, R. e Wehner, R. (1997), "An autonomous agent with a polarized light compass", in *Adaptive Behavior*, 6, pp. 131-161.

Vecchi problemi filosofici per la nuova intelligenza artificiale

- Mataric, M.J. (1990), "Navigating with a rat brain: a neurobiologically-inspired model for robot spatial representation", in *From Animals to Animats*, Cambridge, MA, MIT Press, pp. 169-175.
- Mataric, M.J. (1995), *Integration of representation into goal-driven behavior-based robots*, in L. Steels e R. Brooks (a cura di), *The Artificial Life Route to Artificial Intelligence*, Hillsdale, NJ, Erlbaum, pp. 165-181.
- Newell, A. (1980), "Physical symbol systems", in *Cognitive Science*, 4, pp. 135-183.
- Newell, A. e Simon, H.A. (1976), "Computer science as empirical inquiry: symbols and search", in *Communications of the ACM*, 19, pp. 113-126.
- Nolfi, S. e Floreano, D. (2000), *Evolutionary Robotics*, Cambridge, MA, MIT Press.
- O'Keefe, J.A. e Nadel, L. (1978), *The Hippocampus as a Cognitive Map*, Oxford, Oxford University Press.
- Prescott, T.J. (1994), "Spatial learning and representation in animats", in *From Animals to Animats 3*, Cambridge, MA, MIT Press, pp. 164-173.
- Prescott, T.J. (1996), "Spatial representation for navigation in animats", in *Adaptive Behavior*, 4, pp. 85-123.
- Scarborough, D. e Sternberg S. (eds.), (1998), *An Invitation to Cognitive Science*, vol. 4, Cambridge, MA, MIT Press.
- Scheibel, A.B. e Schopf, J.W. (eds), (1997), *The Origin and Evolution of Intelligence*, Boston, Jones and Bartlett.
- Simon, H.A. (1973), "The structure of ill structured problems", in *Artificial Intelligence*, 4, pp. 181-201.
- Steels, L. e Brooks, R. (eds.), (1995), *The Artificial Life route to Artificial Intelligence*, Hillsdale, NJ, Erlbaum
- Tamburrini, G. (2002), *I matematici e le macchine intelligenti*, Milano, Bruno Mondadori.
- Trautteur, G. (ed.), (1995), *Consciousness: Distinction and Reflection*, Napoli, Bibliopolis.
- Trautteur, G. (2002), "Undici tesi sulla scienza cognitiva", in *Adelphiana*, 1, pp. 71-96, www.adelphiana.it
- van Gelder, T. (1995), "What might cognition be, if not computation", in *Journal of Philosophy*, 91, pp. 345-381.
- VanLehn, K. (ed.), (1991), *Architecture for Intelligence*, Hillsdale, NJ., Erlbaum.
- Vera, A.H. e Simon, H.A. (1993a), "Situated action: a symbolic interpretation", in *Cognitive Science*, 17 pp.7-48.
- Vera, A.H. e Simon, H.A. (1993b), "Situated action: a reply to William Clancey", in *Cognitive Science*, 17, pp. 117-133.
- Vera, A.H. e Simon, H.A. (1994), "Reply to Touretzky and Pomerleau: reconstructing PSSs", in *Cognitive Science*, 18, pp. 355-360.
- Walter, W. G. (1953), *The Living Brain*, London, Duckworth (trad. it. *Il cervello vivente*, Milano, Feltrinelli, pp. 1957).
- Wehner, R. (1997a), *Preparational intelligence: how insects and birds find their way*, in Scheibel e Schopf 1997, pp. 1-26.
- Wehner, R. (1997b), "Sensory systems and behaviour", in Krebs e Davies (1997), pp. 19-41.
- Wehner, R., Michel, B. e Antonsen, P. (1996), "Visual navigation in insects: coupling of egocentric and geocentric information", in *Journal of Experimental Biology*, 199, pp. 129-140.
- Winograd, T. e Flores, F. (1986), *Understanding Computers and Cognition*, Norwood, NJ, Ablex (trad. it. *Calcolatori e conoscenza*, Milano, Mondadori, 1987).
- Wyeth, G. e Browning, B. (1998), "Cognitive models of spatial navigation from a robot builder's perspective", in *Adaptive Behavior*, 6, pp. 509-534.