

Verso una scienza epistemologica del senso comune.

Il progetto scientifico di John McCarthy*

Fausto Giunchiglia^{1,2} Paolo Bouquet^{1,3}

¹D.I.S.A. – Università di Trento

²Istituto per la Ricerca Scientifica e Tecnologica – ITC (Trento)

³Dipartimento di Filosofia – Università di Genova

fausto@cs.unitn.it bouquet@cs.unitn.it

1 Introduzione

L'obiettivo di questo nostro contributo è quello di presentare e discutere un approccio all'Intelligenza Artificiale (IA) che è stato proposto da John McCarthy alla fine degli anni '50 e che pone al centro della ricerca la nozione di *senso comune* (*common sense*). Riteniamo infatti che tale approccio sia non soltanto uno dei tentativi più organici di dare solide basi alla ricerca in IA, ma anche che esso introduca una serie di concetti e ponga una serie di problemi che possono risultare di estremo interesse per una riflessione filosofica sulle nozioni di normatività logica e ragionamento di senso comune.

Uno degli obiettivi di lungo periodo dell'IA è quello di costruire delle macchine (dei programmi di computer) che esibiscano dei comportamenti che in esseri umani chiameremmo intelligenti. Questo obiettivo generale si è storicamente articolato in molti filoni di ricerca, che comprendono tra gli altri lo studio della visione artificiale, della sintesi e del riconoscimento vocale, della comprensione del linguaggio naturale, della rappresentazione della

*Questo lavoro ha beneficiato dei commenti e dei suggerimenti, tra gli altri, di Vladimir Lifschitz, John McCarthy, Carlo Penco, Manuela Viezzer e Toby Walsh.

conoscenza e del ragionamento. È proprio in quest'ultimo filone che si colloca la maggior parte del lavoro di McCarthy. Esso si basa sulla cosiddetta ipotesi della rappresentazione della conoscenza [Brachman & Levesque 1985], ovvero sull'ipotesi che un'entità, per poter esser detta intelligente (qualunque cosa questo significhi), debba possedere una certa quantità di conoscenza sul proprio ambiente e debba essere capace di usarla per ricavarne di nuova e per orientare il proprio comportamento. L'idea di McCarthy che qui intendiamo discutere è che il successo di questo programma dipende in modo cruciale dalla capacità di dotare i programmi di computer di quelle conoscenze e di quelle abilità di ragionamento che in un essere umano sarebbero chiamate di *senso comune*. Questa idea è stata presentata e discussa in diverse fasi del lavoro di McCarthy e quasi mai in modo organico. Il nostro lavoro sarà quello di far emergere dalle osservazioni, dalle riflessioni e dai contributi scientifici di McCarthy le linee generali di un progetto di ricerca di lungo termine e che, per le ragioni che spiegheremo, ci pare adeguato chiamare un'*epistemologia scientifica del senso comune*². In particolare, discuteremo in maniera critica i seguenti aspetti:

- cosa significa fare un'epistemologia del senso comune?
- quali strumenti sono necessari per fare di questa epistemologia una scienza del senso comune?
- quali sono i problemi che McCarthy (e altri dopo di lui) hanno affrontato di questa scienza? Quali soluzioni sono state proposte?
- quali sono i problemi ancora aperti?

Concluderemo con alcune riflessioni sulla portata e sul futuro di questo progetto.

¹Nel testo, useremo la parola "epistemologia" come traduzione dell'Inglese "epistemology" con il significato di teoria della conoscenza e non di filosofia della scienza come talvolta è usata in Italiano.

²La caratterizzazione del progetto di McCarthy come "epistemologia scientifica del senso comune" è stata introdotta per la prima volta in [Giunchiglia 1995].

2 Epistemologia del senso comune

2.1 Senso comune

La prima formulazione esplicita del programma di ricerca di lungo termine di McCarthy si trova nell'articolo [McCarthy 1959]³. Ecco come McCarthy stesso identifica le caratteristiche che un programma (da lui chiamato *advice taker*) dotato di senso comune dovrebbe possedere:

The *advice taker* is a proposed program for solving problems by manipulating sentences in formal languages. ...

The main advantages that we expect the *advice taker* to have is that its behavior will be improvable merely by making statements to it, telling it about its symbolic environment and what is wanted from it. To make these statements will require little if any knowledge of the program or the previous knowledge of the *advice taker*. One will be able to assume that the *advice taker* will have available to it a fairly wide class of immediate consequences of anything it is told and its previous knowledge. This property is expected to have much in common with what makes us describe certain humans as having *common sense*. We shall therefore say that *a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.*

La citazione appena riportata suggerisce quelle che, a parere di McCarthy, sono le tre proprietà che permettono di stabilire se un programma di computer possiede senso comune:

Proprietà 1 innanzitutto, deve avere conoscenza di un numero sufficientemente elevato di fatti a proposito dell'ambiente in cui vive;

Proprietà 2: in secondo luogo, deve poter incrementare questa conoscenza deducendone un insieme sufficientemente vasto di conseguenze immediate;

³Tutti gli articoli di McCarthy che citeremo nel corso dell'articolo sono stati recentemente ristampati in [Lifschitz 1990]. Le citazioni sono tutte tratte da questa ristampa.

Proprietà 3: infine, deve essere in grado di trasformare in nuova conoscenza ciò che gli viene detto.

La terza di queste proprietà è cruciale nel lavoro di McCarthy e concerne la capacità di comunicare a un computer quello che gli essere umani si comunicano mediante l'uso del linguaggio ordinario: "What we can learn from natural language is not how to express in English what we already know how to express in computerese. Rather, we must study those ideas expressible in natural language that no one knows how to represent at all in a computer". Le prime due proprietà, invece, pongono le fondamenta di una distinzione molto importante per McCarthy, ovvero la distinzione tra ciò che l'*advice taker* sa e la capacità di dedurre nuovi fatti dall'informazione già in suo possesso. La distinzione tra l'insieme di fatti che possiamo considerare *conoscenza di senso comune* di un programma e la capacità di inferire nuovi fatti mediante forme di *ragionamento di senso comune*, è posta alla base di una fondamentale distinzione fatta in [McCarthy & Hayes 1969] tra due parti della ricerca in IA, la parte *epistemologica* e la parte *euristica*:

The epistemological part is the representation of the world in such a form that the solution of problems follows from the facts expressed in the representation. The heuristic part is the mechanism that on the basis of the information solves the problem and decides what to do.

La critica che McCarthy, nel 1969, rivolgeva alla comunità dell'IA è quella di essersi concentrata in modo prioritario e prevalente sulla parte euristica, trascurando quella epistemologica o studiandola quel tanto che era ritenuto necessario per migliorare l'efficienza nella soluzione automatica di problemi. Per contro, la ricerca di McCarthy è fondamentale rivolta al problema epistemologico. [McCarthy 1977] motiva questa scelta sostenendo che una soluzione soddisfacente al problema epistemologico può essere poi utilizzata per una varietà di soluzioni al problema euristico, dato che una buona rappresentazione è la base per qualunque soluzione che voglia aspirare a un elevato grado di generalità. Vediamo meglio in che senso la parte epistemologica dell'IA può essere vista fondamentale come un problema di adeguatezza rappresentativa.

2.2 Adeguatezza epistemologica

Il primo passo di McCarthy nello studio della parte epistemologica dell'IA è la classificazione dei (possibili) formalismi di rappresentazione in base a tre nozioni di adeguatezza. Ecco come la classificazione è introdotta in [McCarthy & Hayes 1969]:

A representation is called *metaphysically adequate* if the world could have that form without contradicting the facts of the aspect of reality that interests us. ...

A representation is called *epistemologically adequate* for a person or machine if it can be used practically to express the facts that one actually has about the aspect of the world. ...

A representation is called *heuristically adequate* if the reasoning processes actually gone through in solving a problem are expressible in the language. [pages 27-28]

Secondo McCarthy, la realizzazione di programmi dotati di senso comune passa inevitabilmente attraverso le idee di adeguatezza epistemologica e di adeguatezza euristica. Curiosamente, McCarthy non sviluppa in dettaglio la nozione di adeguatezza euristica, forse perchè in realtà questa nozione può essere fatta collapsare su quella di adeguatezza epistemologica. Infatti, a nostro avviso, chiedere che una rappresentazione sia euristicamente adeguata significa sostanzialmente richiedere che essa sia una rappresentazione epistemologicamente adeguata del ragionamento di senso comune. Se così stanno le cose, l'adeguatezza euristica verrebbe a essere una nozione derivata e, in definitiva, meno importante di quella di adeguatezza epistemologica. Per quanto riguarda invece la nozione di adeguatezza metafisica, McCarthy ritiene che essa sia un requisito fondamentale per ogni teoria scientifica, ma che da sola non sia sufficiente per dotare un'entità artificiale di senso comune: "The skeptic who doubts whether there is anything to say about the world apart from the particular sciences should try to write a computer program that can figure out how to get to Timbuktoo, taking into account, not only the facts about travel in general, but also the facts about what people and documents have what information, and what information will be required at different stages of the trip and when and how it is to be obtained . He will rapidly discover that he ... will be unable to express and build into a program 'what everybody knows' " [McCarthy 1979a].

Concentriamoci dunque sulla nozione cruciale di adeguatezza epistemologica. In un certo senso, possiamo dire che la maggior parte del lavoro tecnico prodotto da McCarthy in articoli ben noti (per esempio [McCarthy 1977, McCarthy 1979a, McCarthy 1979b, McCarthy 1980, McCarthy 1986, McCarthy 1987]) ha lo scopo di sviluppare formalismi di rappresentazione epistemologicamente adeguati. Al di là delle differenze puntuali, è chiaro che l'idea di adeguatezza epistemologica pone due vincoli metodologici su qualsiasi possibile rappresentazione, entrambi impliciti nella parola *practically* che compare nel brano riportato poco sopra. Il primo vincolo è di tipo ontologico e ha a che fare con gli “oggetti” che popolano il mondo del senso comune: il linguaggio di rappresentazione prescelto deve permettere di rappresentare l'informazione disponibile a un programma allo stesso livello di astrazione usato dagli esseri umani nelle loro attività di ogni giorno. Questo vincolo ha lo scopo di permettere a un programma di interagire con gli esseri umani senza dover costruire dei meccanismi di comunicazione *ad hoc*. Infatti, non è possibile istruire un programma dandogli informazioni che esso non è nemmeno in grado di rappresentare. Per esempio, non è chiaro come un programma possa apprendere che John crede che Espero e Fosforo sono lo stesso pianeta se la conoscenza di tale programma è rappresentata in base al linguaggio della fisica delle particelle o della meccanica quantistica (due esempi – secondo McCarthy – di linguaggi di rappresentazione metafisicamente adeguati, ma epistemologicamente inadeguati).

Il secondo vincolo è di tipo teorico: le teorie che sviluppiamo devono avere una forma tale da permettere all'*advice taker* di acquisire “abbastanza” conoscenza di senso comune e di eseguire “abbastanza” ragionamento di senso comune. Più concretamente, questo significa che una rappresentazione epistemologicamente adeguata deve permettere al computer di rappresentare “the information actually available to a subject under given circumstances”, laddove una rappresentazione inadeguata non gli permetterebbe di “achieve enough goals requiring general intelligence no matter how fast [it was] allowed to run”. In breve, non solo uno schema di rappresentazione deve essere basato su un'ontologia corretta; esso deve anche usare teorie del mondo che di fatto rappresentano l'informazione effettivamente disponibile in una determinata situazione in modo tale da permettere all'*advice taker* di inferire ciò che gli serve per raggiungere un determinato obiettivo.

Dal nostro punto di vista, l'adeguatezza epistemologica è la nozione più importante introdotta da McCarthy a livello metodologico ed è assolutamente centrale nel suo programma di ricerca. Essa, infatti, definisce implicitamente

il tipo di formalismi e di teorie del senso comune che devono essere sviluppate e, quindi, elimina un gran numero di possibili opzioni. Per esempio, il concetto di adeguatezza epistemologica distingue l'IA dalla neuro-fisiologia e da altri approcci a livello biologico (si tenga presente che questo punto è stato – ed è – piuttosto controverso. Cfr. [Minsky 1985, Brooks 1991]). Inoltre, il concetto di adeguatezza epistemologica permette di stabilire le relazioni tra IA e psicologia. Infatti, McCarthy ritiene che l'oggetto della psicologia sia lo studio di come certi processi (per esempio, il ragionamento) sono realizzati negli esseri umani, mentre l'IA ha a che fare con metodi generali per risolvere problemi, indipendentemente dal fatto che chi li risolve sia un essere umano, un Marziano o un programma di computer. La richiesta di rispettare l'adeguatezza epistemologica produce (e, di fatto, ha storicamente prodotto) lo sviluppo di teorie del senso comune al livello che [Newell 1982] definisce al livello della conoscenza (*knowledge level*). Questo è ciò che chiamiamo *un'epistemologia del senso comune*.

3 Una scienza del senso comune

McCarthy è interessato a quella che potremmo chiamare una *scienza del senso comune*, cioè allo studio dei modi in cui il senso comune si manifesta in natura e nella formulazione di leggi che lo descrivano in termini generali. In molti articoli, McCarthy discute in che tipo di scienza debba svilupparsi l'IA. In particolare, questo problema è discusso in relazione a due altre discipline, vale a dire la logica e la filosofia, ovvero le discipline che a suo parere hanno più in comune con l'IA (o almeno con la sua parte epistemologica). Cerchiamo di vedere qual è la sua posizione.

3.1 IA e filosofia

McCarthy ritorna più volte sui rapporti tra IA e filosofia e sulla rilevanza che l'una può avere per l'altra. Se scorriamo ciò che egli ha scritto su questo argomento, è possibile descrivere la sua posizione in merito in tre punti. Prima di tutto, McCarthy ritiene che IA e filosofia condividano l'interesse per l'epistemologia. Infatti, egli ritiene che realizzare un programma di computer richieda necessariamente di dotarlo di una visione del mondo generale, indipendentemente dai contenuti che gli verranno trasmessi dalle varie scienze particolari. In secondo luogo, se è vero che vi è in interesse comune per

l'epistemologia, McCarthy ritiene che le prospettive da cui essa è studiata siano diverse in IA e in filosofia: mentre in filosofia l'accento è posto su ciò che un agente può conoscere in condizioni ottimali e senza alcun vincolo sulle sue risorse osservative e computazionali, in IA l'accento è posto invece su ciò che un agente può conoscere date determinate risorse osservative e di calcolo. In terzo luogo, McCarthy ritiene che molti dibattiti di tipo filosofico possano essere esclusi da una scienza del senso comune in quanto incapaci di fornire alcuna indicazione concreta di come costruire l'*advice taker*: "Since the philosophers have not really come to an agreement in 2500 years, it might seem that artificial intelligence is in a rather hopeless state if it is to depend on getting concrete enough information out of philosophy to write computer programs. Fortunately, merely undertaking to embody the philosophy in a computer program involves making enough philosophical presuppositions to exclude most philosophy as irrelevant" [McCarthy & Hayes 1969]. L'idea è che qualunque distinzione concettuale di tipo filosofico (per esempio quella tra realismo ed empirismo) è importante (per l'IA) solo in quanto è in grado di suggerire modi alternativi di costruire l'*advice taker*; se così non è, la distinzione si rivela ininfluenza.

3.2 Il ruolo della logica

Le proprietà 2 e 3 richiedono che l'*advice taker* sia in grado di accrescere la propria conoscenza in modo sensato e indipendentemente dal proprio stato corrente. La proprietà 2 richiede che questo sia ottenuto mediante un processo di ragionamento; la proprietà 3 richiede che questo sia ottenuto mediante l'interazione con il mondo esterno. Come descritto in [McCarthy 1959], un modo molto vantaggioso di ottenere questo risultato è quello di usare dei linguaggi di rappresentazione dichiarativi, in opposizione a quelli che McCarthy chiama linguaggi imperativi (e che, in seguito, furono chiamati procedurali). Tra i linguaggi dichiarativi, McCarthy ritiene che la scelta migliore sia di usare linguaggi logici [McCarthy 1959, McCarthy & Hayes 1969, McCarthy 1988]. Essi infatti, a differenza di altri linguaggi, offrono una nozione formale di deduzione e un insieme di strumenti che permettono di studiare le proprietà dei formalismi sviluppati, come per esempio la loro espressività e la loro consistenza.

Da questo punto di vista, la logica diviene più uno strumento della scienza del senso comune che un oggetto di interesse in sé e per sé; pertanto, essa dovrebbe essere sviluppata nella misura in cui essa è necessaria per lo sviluppo

dell'*advice taker*. Questo punto è così espresso da McCarthy in [McCarthy 1983a]:

AI badly needs mathematical and logical theory, but the theory required involves conceptual innovations — not just mathematics. We won't reach human level intelligence by more algorithms reducing the complexity of a problem from n^2 to $n \log n$, and still less by proofs that yet another problem is unsolvable or NP-complete. Of course, these results are often very significant as mathematics or computer science.

Come è ben noto, la scelta di un approccio all'IA basato sulla logica ha incontrato forti resistenze (si vedano per esempio [Minsky 1985, Brooks 1991]). Il problema principale è che la logica è stata sviluppata con obiettivi alquanto diversi da quelli dell'IA (per esempio, per dimostrare la consistenza del ragionamento matematico o per dare una semantica a porzioni del linguaggio naturale). Di conseguenza, non dovrebbe stupire che la logica, per quanto sia un ottimo punto di partenza per una scienza del senso comune, non sia sufficientemente espressiva per rappresentare tutto ciò che appartiene al senso comune. Ecco perchè molti dei lavori più tecnici di McCarthy sono volti a sviluppare alcune delle estensioni di cui la logica ha bisogno per raggiungere il livello dell'adeguatezza epistemologica.

4 Problemi

Il lavoro discusso nei paragrafi precedenti fissa l'orizzonte per la ricerca che è necessario fare. Il passo successivo, discusso da McCarthy in lavori quali [McCarthy 1977, McCarthy 1984], è di isolare i problemi rilevanti da affrontare.

Per quanto riguarda la conoscenza di senso comune, McCarthy individua tre argomenti principali. Il primo concerne la conoscenza di senso comune sul tempo. Con ciò, McCarthy intende essenzialmente la formalizzazione del ragionamento sul modo in cui le nostre azioni (o altri eventi) influenzano l'evoluzione del mondo lungo la dimensione del tempo. Per esempio, l'*advice taker* deve essere in grado di ragionare sul fatto che prendere l'aereo da Glasgow a Londra e da Londra a Mosca avrà come effetto quello di trasportare il viaggiatore a Mosca e non, per esempio, quello di far cambiare il colore dei vestiti del viaggiatore stesso (normalmente). Il secondo argomento concerne

la conoscenza di senso comune sullo spazio. Gli oggetti del senso comune sono immersi nello spazio; spesso cambiano posizione e talvolta forma. Tutto ciò è di fondamentale importanza per un robot mobile. Il terzo argomento è la conoscenza di senso comune sugli stati interni. Il mondo è popolato di oggetti (persone, programmi, termostati, ecc.) che hanno stati interni i quali influenzano il loro comportamento. Per interagire (cooperare, competere, ecc.) con tali oggetti, è molto utile essere in grado di rappresentare e di ragionare sui loro stati interni. Un tipico esempio è quello di ragionare sulle credenze (aspettative, timori, intenzioni) di altri agenti al fine di prevederne i comportamenti.

McCarthy discute anche un certo numero di problemi che sorgono nella formalizzazione del ragionamento di senso comune. Molti di questi problemi sono casi particolari di un unico problema, quello cioè dell'abilità degli esseri umani di "saltare alle conclusioni". Un tipico esempio di problema di questo tipo (e forse il più interessante) è il cosiddetto *problema della qualificazione*, discusso in particolare in [McCarthy & Hayes 1969, McCarthy 1977]. McCarthy discute il problema soprattutto in relazione al ragionamento su azioni, ma in termini generali si tratta dell'impossibilità di tenere effettivamente conto di tutti i possibili fatti che qualificano la verità di un enunciato di senso comune⁴. Nel caso delle azioni, per esempio, il problema si manifesta nell'impossibilità pratica (e probabilmente anche teorica) di ragionare su tutte le possibili precondizioni che devono essere soddisfatte affinché un'azione sia eseguibile. Un problema analogo è il ben noto *frame problem*, ovvero il problema di rappresentare l'esecuzione di un'azione in modo tale da poter ragionare su ciò che rimane immutato dopo la sua esecuzione. Infine, vale la pena di ricordare che, per McCarthy, attribuire qualità mentali a macchine o persone significa di fatto saltare alla conclusione che quella macchina o quella persona abbia effettivamente certe credenze [McCarthy 1979a, McCarthy 1983b].

Ciò non significa che il ragionamento di senso comune si riduca alla capacità di saltare alle conclusioni. Tuttavia, come dimostrano lavori quali [McCarthy 1959, McCarthy 1979b, McCarthy 1990], molte altre forme di ragionamento di senso comune possono essere ricondotte alla logica deduttiva classica, posto che si trovi un modo adeguato di rappresentare la conoscenza necessaria. Invece, il processo di saltare alle conclusioni non è in linea di principio riconducibile al ragionamento deduttivo a causa della sua non

⁴Questa generalizzazione del problema è ampiamente discussa in [Bouquet & Giunchiglia 1995, Giunchiglia, Giunchiglia, Costello & Bouquet 1996].

monotonicità. La logica classica è per definizione monotona, cioè per essa vale la seguente proprietà: se $\Gamma \vdash A$ e $\Gamma \subseteq \Gamma'$, allora $\Gamma' \vdash A$, dove A e B sono formule e Γ, Γ' sono insiemi di formule. Ma questo non vale nel processo di saltare alle conclusioni. Il tipico esempio è quello di Tweety. Dal fatto che Tweety è un uccello è in genere ovvio inferire che Tweety vola; tuttavia, qualora aggiungessimo il fatto che Tweety è un pinguino, tale conclusione dovrebbe venir ritirata. Meccanismi analoghi di ragionamento “fallibile” sono per McCarthy necessari per formalizzare convenzioni comunicative tra agenti, informazioni di tipo probabilistico non numerico, leggi della fisica di senso comune, ecc.

5 Soluzioni proposte

Nell’ambito di quello che McCarthy chiama senso comune, egli non si interessa in particolare alla rappresentazione dell’informazione spaziale. Per contro, la maggior parte del suo lavoro formale è volto a proporre soluzioni al problema della rappresentazione del ragionamento su azioni, degli stati interni e dell’abilità della gente di saltare alle conclusioni. Noi riassumeremo le linee principali di questo lavoro.

5.1 Conoscenza su tempo e azioni

Per rappresentare la conoscenza e il ragionamento di senso comune su tempo e azioni McCarthy (assieme a Pat Hayes) introduce il cosiddetto *situation calculus* [McCarthy & Hayes 1969]. Il nome di questo formalismo deriva dal fatto che esso include, come parte della propria ontologia di base, un nuovo tipo di oggetto, vale a dire le *situazioni*, dove una situazione è definita come lo stato completo dell’universo in un determinato istante di tempo. L’inclusione delle situazioni nel dominio del calcolo permette di usarle come argomento di simboli funzionali e predicativi. Tecnicamente, McCarthy chiama *fluenti* tutti quei simboli il cui dominio include lo spazio delle situazioni. In particolare, un fluente si dice proposizionale se ha come valore un valore di verità, mentre si dice situazionale se ha come valore una situazione. Ecco alcuni esempi di fluenti discussi da McCarthy e Hayes:

- $at(p, x, s)$ (l’agente p si trova nel punto x nella situazione s) e $in(x, y, s)$ (l’oggetto x è all’interno del luogo y nella situazione s). I due fluen-

ti permettono di esprimere alcuni semplici fatti sulla posizione degli oggetti e sulle loro relazioni spaziali.

- $time(s)$ (il tempo della situazione s). Questo fluente permette di esprimere proposizioni sul tempo a cui per esempio un certo fatto è vero o falso.
- $result(p, \sigma, s)$ (la situazione che risulta dal fatto che la persona p esegue l'azione σ nella situazione s). Questo importante fluente permette di esprimere proposizioni sull'evoluzione temporale del mondo in seguito all'esecuzione di certe azioni da parte di agenti.

McCarthy non dà molte motivazioni per l'introduzione del situation calculus. In particolare, è interessante notare che egli non sente il bisogno di giustificare la scelta di un calcolo basato sulle situazioni piuttosto che sugli istanti di tempo. McCarthy accenna al fatto che un calcolo basato sulle situazioni ha la stessa capacità espressiva e inferenziale di un calcolo basato sul tempo. In effetti, di primo acchito la scelta di fare una logica temporale sembra quella più naturale. Tuttavia, è probabile che la scelta di McCarthy sia dettata dall'idea di adeguatezza epistemologica. Infatti, la maggior parte del ragionamento di senso comune sulle azioni non concerne tanto il tempo, quanto piuttosto gli effetti che una certa azione produce sulla situazione di partenza. Solo in certi casi viene esplicitamente introdotto il tempo (per esempio, per azioni come prendere l'aereo, cioè azioni che hanno come preconditione il fatto di essere eseguite a un preciso istante di tempo) ed è proprio ciò che il fluente $time(s)$ permette di fare nel situation calculus.

5.2 Conoscenza sugli stati interni

[McCarthy 1979a] propone di formalizzare il ragionamento sugli stati interni nei termini di alcune primitive astratte di alto livello dette *qualità mentali*, tra le quali McCarthy include per esempio credenze, conoscenze, coscienza (e auto-coscienza), desideri, intenzioni. Secondo McCarthy, questa scelta è necessaria se vogliamo avere teorie epistemologicamente adeguate delle macchine, specialmente quando esse sono molto complesse e non completamente conosciute. In [McCarthy 1983b], troviamo la seguente formulazione di quanto appena detto: “the language [used to describe internal states] must be able to express the information our program can actually get about a person’s or machine’s ‘state of mind’ — not just what might be obtainable

if the neurophysiology of the human or the design of the machine were more accessible”.

In modo un po' impreciso, possiamo dire che le qualità mentali di cui parla McCarthy sono un soprainsieme di quelli che sono comunemente chiamati atteggiamenti proposizionali. Storicamente, le formalizzazioni più usate degli atteggiamenti proposizionali sono basate sulla logica modale (cfr. per esempio [Hintikka 1962]). McCarthy non crede invece che le logiche modali siano uno strumento adeguato. In [McCarthy 1979b] egli difende l'uso della logica del primo ordine per vari motivi. Innanzitutto, molti dei problemi che si presentano nella formalizzazione del senso comune possono essere risolti usando la logica del primo ordine; in secondo luogo, non esistono dei buoni dimostratori automatici per la logica modale, mentre ne esistono di molto potenti per la logica del primo ordine; in terzo luogo, molti lavori sulla dimostrazione automatica di teoremi in logica modale (per esempio [Morgan 1976]) si basano sostanzialmente su una traduzione della logica modale in una logica del primo ordine; infine, l'adozione della logica modale per la formalizzazione delle qualità mentali costringe a modificare sia la sintassi, sia la semantica della logica del primo ordine, costringendo poi ad applicare questa diversa sintassi e semantica anche ad altre porzioni del senso comune per la quali sarebbe molto più semplice mantenere quella del primo ordine.

Di fronte a queste obiezioni, McCarthy propone un formalismo per ragionare sulle qualità mentali. La sua proposta si basa su due idee fondamentali:

- la prima è quella di reificare concetti e proposizioni come oggetti del linguaggio di rappresentazione;
- la seconda è quella di reificare i mondi possibili come oggetti del linguaggio di rappresentazione.

La prima idea gli permette di ottenere la stessa espressività della logica modale proposizionale senza averne gli svantaggi: “Admitting individual concepts as objects — with concept-valued constants, variables, functions, and expressions — allows ordinary first order theories of necessity, knowledge, belief, and wanting without modal operators or quotation marks and without the restrictions on substituting equals for equals that either device makes necessary. ... [This formalization] doesn't modify the logic and is more powerful, because it includes mappings from objects to concepts” [McCarthy 1979b].

La seconda idea gli permette di ottenere la stessa potenza espressiva della logica modale quantificata, anche se un risultato tecnico di equivalenza non è mai stato effettivamente fornito. Tuttavia, è importante notare che McCarthy si è sempre dimostrato abbastanza riluttante verso l'accettazione di questo secondo processo di reificazione, ritenendo la stessa idea di mondo possibile non del tutto adeguata alla formalizzazione del senso comune.

5.3 Saltare alle conclusioni

Il formalismo proposto da McCarthy per formalizzare il ragionamento non monotono è quello di *circumscription*⁵. Esso si basa su una formula che, aggiunta a un certo insieme di fatti Γ , permette di saltare alla conclusione che gli oggetti per i quali è possibile dimostrare che possiedono una certa proprietà P (usando i fatti contenuti in Γ) sono anche i soli oggetti che possiedono la proprietà P . La formula che esprime la circumscription è stata presentata da McCarthy in molti modi diversi [McCarthy 1977, McCarthy 1980, McCarthy 1986]. Una delle versioni più intuitive è la seguente. Sia P un simbolo di predicato e A un enunciato di un linguaggio del primo ordine. Indichiamo con $A(\Phi)$ il risultato di sostituire tutte le occorrenze di P in A con Φ . Allora, la circumscription di P in A è lo schema di enunciati:

$$A(\Phi) \wedge \forall \bar{x} (\Phi(\bar{x}) \rightarrow P(\bar{x})) \rightarrow \forall \bar{x}.(P(\bar{x}) \rightarrow \Phi(\bar{x})) \quad (1)$$

dove \bar{x} sta per la n-pla x_1, \dots, x_n . Intuitivamente, questa formula dice che se vale $A(\Phi)$ e se Φ ha un'estensione minore di P , allora P e Φ hanno la stessa estensione. In altre parole, l'insieme di oggetti che soddisfano P è reso tanto piccolo quanto quello degli oggetti che soddisfano Φ . Così, per esempio, sia P il predicato *isblock*(x) (x è un blocco) e A sia la seguente formula:

$$isblock(B) \wedge isblock(C) \wedge isblock(D) \quad (2)$$

Applicando la formula (1) al predicato *isblock* nell'insieme di formule (2) è possibile inferire la seguente formula:

$$\forall x (isblock(x) \supset (x = B \vee x = C \vee x = D))$$

che dice che B, C, D sono i soli oggetti che sono blocchi.

⁵Circumscription, come il situation calculus, è uno dei formalismi più conosciuti e studiati in IA. Chi fosse interessato può trovare una buona quantità di materiale su circumscription (e su altre logiche non monotone) in [Ginsberg 1987, Lifschitz 1993].

6 Risultati e problemi aperti

A partire dalla formulazione del progetto dell'*advice taker* nel 1959, la ricerca di McCarthy in IA ha percorso tre principali passi concettuali:

1. innanzitutto, l'identificazione della necessità di dotare un programma intelligente di senso comune e dunque di studiare quella che abbiamo chiamato un'epistemologia del senso comune. In quest'ottica, possiamo probabilmente dire che i maggiori risultati del lavoro di McCarthy sono stati l'identificazione del senso comune come l'aspetto chiave della descrizione dell'*advice taker* e la nozione di adeguatezza epistemologica;
2. in secondo luogo, l'identificazione della necessità di una scienza del senso comune. Qui il maggior risultato è probabilmente la proposta di un approccio all'IA basato sulla logica;
3. infine, lo sviluppo di alcune proposte concrete. In particolare, l'identificazione di alcuni importanti problemi epistemologici da risolvere, lo sviluppo di alcune soluzioni tecniche, il riconoscimento di alcune difficoltà centrali per la realizzazione del progetto dell'*advice taker*.

In conclusione, si può dire che il progetto di un *advice taker* è ancora ben lungi dall'essere realizzato. Tuttavia, a partire dal 1959, molti progressi sono stati fatti nel solco tracciato da McCarthy e per molte nuove linee di ricerca sono state gettate le basi. Seguendo le linee proposte da McCarthy, l'IA ha trovato un proprio chiaro obiettivo (lo studio del senso comune), un metodo (l'epistemologia del senso comune) e uno strumento (la logica formale). Questa "maturazione" dell'IA è in buona parte dovuta a McCarthy, che ha tracciato le linee di quella che abbiamo chiamato una scienza epistemologica del senso comune. Questo, ci pare, è il maggior risultato del lavoro di McCarthy, le cui idee in proposito non solo possono ancora guidare la ricerca in IA e nell'informatica teorica (basti pensare alle recenti riflessioni di McCarthy sul legame tra problema della generalità e contesto), ma anche stimolare nuove riflessioni e produrre nuovi approcci in discipline quali la filosofia e le scienze cognitive.

Riferimenti bibliografici

- Bouquet, P. & Giunchiglia, F. [1995]. Reasoning about Theory Adequacy. A New Solution to the Qualification Problem, *Fundamenta Informaticae* **23**(2-4): 247-262. Also IRST-Technical Report 9406-13, IRST, Trento, Italy.
- Brachman, R. & Levesque, H. [1985]. *Readings in knowledge representation*, Morgan Kaufmann Publishers, Inc., 95, First Street, Los Altos, CA 94022.
- Brooks, R. A. [1991]. Intelligence Without Reasoning, *Proceedings IJCAI-91, 12th Int. Joint Conf. on Artificial Intelligence*, Sydney, Australia, pp. 569-595.
- Ginsberg, M. (ed.) [1987]. *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann.
- Giunchiglia, F. [1995]. An epistemological science of common sense, *Artificial Intelligence* **77**(2): 371-392. Also IRST-Technical Report 9503-09, IRST, Trento, Italy.
- Giunchiglia, F., Giunchiglia, E., Costello, T. & Bouquet, P. [1996]. Dealing with Expected and Unexpected Obstacles, *Journal of Experimental and Theoretical Artificial Intelligence* **8**. Also IRST-Technical Report 9211-06, IRST, Trento, Italy.
- Hintikka, J. . [1962]. *Knowledge and Belief*, Cornell University Press, Ithaca, NY.
- Lifschitz, V. [1993]. Circumscription, in D. Gabbay, C. Hogger & J. Robinson (eds), *The Handbook of Logic in AI and Logic Programming*, Vol. 3, Oxford University Press, pp. 298-352.
- Lifschitz, V. (ed.) [1990]. *Formalizing Common Sense - Papers by John McCarthy*, Ablex Publishing Corporation.
- McCarthy, J. [1959]. Programs with Common Sense, *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, London: Her Majesty's Stationery Office, pp. 75-91. Also in V. Lifschitz (ed.),

- Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 9–20.
- McCarthy, J. [1977]. Epistemological Problems of Artificial Intelligence, *Proc. of the 5th International Joint Conference on Artificial Intelligence*, pp. 1038–1044. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 77–92.
- McCarthy, J. [1979a]. Ascribing Mental Qualities to Machines, in M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence*, Humanities Press, pp. 161–195. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 93–118.
- McCarthy, J. [1979b]. First Order Theories of Individual Concepts and Propositions, in J. Hayes, D. Michie & L. Mikulich (eds), *Machine Intelligence 9*, Ellis Horwood, pp. 129–148. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 119–141.
- McCarthy, J. [1980]. Circumscription - A Form of Non-monotonic Reasoning, *Artificial Intelligence* **13**(1,2): 27–39. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 142–157.
- McCarthy, J. [1983a]. AI Needs more Emphasis on Basic Research, *AI Magazine* **4**. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 187–188.
- McCarthy, J. [1983b]. The Little Thoughts of Thinking Machines, *Psychology Today* **17**(12). Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 179–186.
- McCarthy, J. [1984]. Some Expert Systems Need Common Sense, in H. Pagels (ed.), *Computer Culture: The Scientific, Intellectual and Social Impact of the Computer*, Vol. 426, Annals of the New York Academy of Sciences, pp. 129–137. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 189–197.
- McCarthy, J. [1986]. Applications of Circumscription to Formalizing Common-Sense Knowledge, *Artificial Intelligence* **28**(1): 71–87. Al-

- so in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 198–225.
- McCarthy, J. [1987]. Generality in Artificial Intelligence, *Communications of ACM* **30**(12): 1030–1035. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 226–236.
- McCarthy, J. [1988]. Mathematical Logic in Artificial Intelligence, *Daedalus* **117**(1): 297–311. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 237–250.
- McCarthy, J. [1990]. Formalization of Two Puzzles Involving Knowledge, in V. Lifschitz (ed.), *Formalizing Common Sense - Papers by John McCarthy*, Ablex Publishing Corporation, pp. 158–166.
- McCarthy, J. & Hayes, P. [1969]. Some Philosophical Problems from the Standpoint of Artificial Intelligence, in B. Meltzer & D. Michie (eds), *Machine Intelligence 4*, Edinburgh University Press, pp. 463–502. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 21–63.
- Minsky, M. [1985]. *The Society of Mind*, Simon & Schuster Inc., New York.
- Morgan, C. [1976]. Methods of Automated Theorem Proving in Nonclassical Logics, *IEEE Transactions on Computers* **C-25**(8): 852–862.
- Newell, A. [1982]. The Knowledge Level, *Artificial Intelligence* **18**(1): 87–127.