

3. Logica e Intelligenza Artificiale

3.1 I modelli computazionali simbolici della mente dalle macchine di Turing all'intelligenza artificiale

Per tracciare una storia dello studio della mente nei termini di computazione simbolica, è opportuno partire dal concetto di *Macchina di Turing* (d'ora in poi MT). Storicamente infatti, le MT hanno fornito sia un modello astratto di macchina calcolatrice da cui hanno preso le mosse lo sviluppo dell'informatica teorica e la tecnologia dei calcolatori digitali, sia il modello di mente cui si sono ispirati i filosofi funzionalisti della mente¹. Quando lo sviluppo dell'informatica ha consentito di tentare una verifica su modelli computazionali reali dell'ipotesi della mente come sistema computazionale, queste due tradizioni sono confluite nell'intelligenza artificiale "classica" di tipo simbolico.

Le MT sono state definite dal logico inglese Alan Turing durante gli anni trenta (Turing 1936-37), quando i calcolatori e l'informatica non esistevano ancora. Lo scopo originario di Turing si collocava nell'ambito delle ricerche logiche sui fondamenti della matematica, e consisteva nell'individuare un equivalente formale del concetto intuitivo di algoritmo, cioè di procedura che consenta di risolvere un problema applicando un numero finito di volte in modo deterministico un insieme (finito) di istruzioni. Si trattava di una questione studiata, da differenti punti di vista, da molti logici in quel periodo (oltre Turing, anche Church, Gödel, Kleene, ed altri), e che ha dato origine a quel ramo della logica matematica detto *teoria della computabilità* o della *calcolabilità effettiva*, o anche *teoria della ricorsività*². Turing propose di risolverlo definendo un modello formale del comportamento di un essere umano che esegue un calcolo di tipo algoritmico³. Tale modello è dato sotto forma di una classe di dispositivi computazionali, di macchine calcolatrici astratte, che in seguito furono dette appunto *macchine di Turing*. Le MT sono macchine astratte nel senso che non vengono presi in considerazione quei vincoli che sono fondamentali se si intende progettare una macchina calcolatrice reale (ad esempio le dimensioni della memoria, il tempo di calcolo, e così via), e soprattutto nel senso che esse sono definite a prescindere dalla loro realizzazione fisica (ad esempio, dal tipo di *hardware* utilizzato). Vale a dire, cosa sia una MT dipende esclusivamente dalle relazioni funzionali che esistono fra le sue parti, e non dal fatto di poter essere costruita con particolari dispositivi materiali.

Non è questa la sede per una trattazione dettagliata delle MT⁴. Ci limitiamo a ricordare che le MT operano su dati che consistono di sequenze finite di simboli appartenenti a un determinato alfabeto (che, per ogni MT, deve essere anch'esso finito). I simboli che una macchina elabora sono scritti su di un dispositivo di memoria, il *nastro* della macchina, di capacità potenzialmente illimitata, ma di cui, in ogni fase del calcolo, può essere utilizzata solo una porzione finita.

Ogni MT è "attrezzata" per eseguire un tipo di calcolo specifico. Dispone cioè di una serie di istruzioni tali che, applicate a partire da una certa situazione iniziale, in cui i dati del calcolo (*l'input*) sono scritti sul nastro mediante una codifica opportuna, esse consentano di raggiungere una situazione finale in cui la macchina si ferma con i risultati (*l'output*) scritti sul nastro. Lo scopo delle istruzioni è di modificare i simboli scritti sul nastro, in modo da ottenere *l'output* voluto a partire dai simboli in *input*. Tali istruzioni, ovviamente, devono essere tali da poter essere applicate in modo deterministico. Una volta fissato l'alfabeto, ogni MT è caratterizzata dall'insieme delle istruzioni, che è chiamato *tavola* di quella macchina.

Le MT possono calcolare un grande numero di funzioni. In particolare, una volta fissato un modo per codificare i numeri naturali mediante i simboli di un alfabeto finito, è possibile definire MT che calcolano le usuali funzioni aritmetiche: addizione, sottrazione, moltiplicazione, elevamento a potenza, e così via. Secondo Turing, le funzioni che possono essere calcolate da una MT sono tutte e sole le funzioni calcolabili per mezzo di un algoritmo. Questa identificazione fra funzioni calcolabili per mezzo di una MT e funzioni algoritmiche va usualmente sotto il nome di *Tesi di Church-Turing*, o *Tesi di Church* (Church 1936) (il logico americano Alonzo Church giunse infatti contemporaneamente alle stesse conclusioni di Turing in modo indipendente, utilizzando un formalismo, il λ -calcolo, equivalente alle MT), e costituisce uno degli enunciati fondamentali della teoria logica della computabilità.

L'interesse verso le MT nell'ambito della teoria delle macchine calcolatrici e in informatica risiede innanzi tutto nel fatto che le MT sono un modello del calcolo algoritmico, quindi di un tipo di calcolo che è, in linea di principio, automatizzabile, eseguibile da un dispositivo meccanico. Ogni MT è quindi il modello astratto di un calcolatore - astratto in quanto prescinde da alcuni vincoli di finitezza cui i calcolatori reali devono sottostare; ad esempio, la memoria di una MT (vale a dire il suo nastro) è potenzialmente estendibile all'infinito (anche se, in ogni fase del calcolo, una MT può sempre utilizzarne solo una porzione finita), mentre un calcolatore reale ha sempre precisi limiti di memoria.

¹ Si veda oltre, in questo stesso paragrafo.

² Si vedano ad esempio (Kleene 1952); (Hermes 1961); (Minsky 1967); (Davis 1958); (Rogers 1967).

³ Questo non significa che Turing intendesse proporre una descrizione del processo psicologico del calcolare. Il suo scopo era piuttosto quello di estrarre quelle caratteristiche del comportamento necessarie per giungere a una caratterizzazione generale ed astratta del concetto di calcolo.

⁴ Per una tale trattazione rimandiamo ai testi citati nella nota precedente.

Vi sono altre ragioni che giustificano l'analogia fra MT e moderni calcolatori digitali. Sino ad ora abbiamo considerato MT che sono in grado di effettuare un solo tipo di calcolo, sono cioè dotate di un insieme di istruzioni che consente loro di calcolare una singola funzione (ad esempio l'addizione, o la moltiplicazione). Esiste tuttavia la possibilità di definire una MT, detta *Macchina di Turing Universale* (d'ora in poi MTU), che è in grado di simulare il comportamento di ogni altra MT. Ciò è reso possibile dal fatto che le istruzioni di ogni MT possono essere rappresentate in maniera tale da poter essere scritte sul nastro di una MT. In questo modo, esse possono essere prese in *input* dalla MTU, la quale, a sua volta, le esegue.

Poiché la MTU è in grado di simulare il comportamento di qualsiasi macchina di Turing, allora essa, in virtù della Tesi di Church, è in grado di calcolare qualsiasi funzione che sia calcolabile mediante un algoritmo. Ciò che caratterizza la MTU rispetto alle macchine di Turing usuali è costituito dal fatto di essere una macchina calcolatrice *programmabile*. Mentre infatti le normali macchine di Turing eseguono un solo programma, che è "incorporato" nella tavola delle loro istruzioni, la MTU assume in *input*, oltre ai dati veri e propri, anche il programma che deve eseguire (cioè, la codifica delle istruzioni della MT che deve simulare). Le istruzioni che compongono la tavola della MTU hanno esclusivamente la funzione di consentirle di interpretare e di eseguire il programma ricevuto in *input*. Un'altra caratteristica fondamentale della MTU è dato dal tipo di trattamento riservato ai programmi. La MTU tratta i programmi (cioè la codifica delle istruzioni della MT da simulare) e i dati (*l'input* della MT da simulare) in maniera sostanzialmente omogenea: essi vengono memorizzati sullo stesso supporto (il nastro), rappresentati utilizzando lo stesso alfabeto di simboli ed elaborati in modo simile. Queste caratteristiche sono condivise dagli attuali calcolatori, i quali presentano la struttura che va sotto il nome di *architettura di von Neumann*.

Storicamente, il primo calcolatore con architettura di von Neumann è stato l'*EDVAC* (acronimo di *Electronic Discrete Variable Automatic Computer*), progettato e realizzato dallo stesso von Neumann presso l'università di Princeton fra il 1945 e il 1952. La struttura di un calcolatore di von Neumann è, molto schematicamente, la seguente. Un dispositivo di *input* e un dispositivo di *output* permettono di accedere dall'esterno alla memoria del calcolatore, consentendo, rispettivamente, di inserirvi e di estrarne dei dati. Le informazioni contenute in memoria vengono elaborate da una singola unità di calcolo (detta *CPU - Central Process Unit*), che opera sequenzialmente su di essi. La caratteristica più importante della macchina di von Neumann è costituita dal fatto che sia dati che programmi vengono trattati in modo sostanzialmente omogeneo, ed immagazzinati nella stessa unità di memoria. Così, quando un programma deve essere eseguito, l'unità di calcolo lo reperisce in memoria, e lo applica quindi ai dati, anch'essi conservati in memoria. Questo consente una grande flessibilità al sistema. Ad esempio, poiché dati e programmi sono oggetti di natura omogenea, è possibile costruire programmi che prendano in *input* altri programmi e li elaborino, e che producano programmi in *output*. Queste possibilità sono ampiamente sfruttate negli attuali calcolatori digitali, e da esse deriva gran parte della loro potenza e della loro facilità d'uso (ad esempio, un compilatore o un sistema operativo sono essenzialmente programmi che operano su altri programmi). In questo senso, un calcolatore di von Neumann è analogo alla MTU. Anche la potenza computazionale è la stessa, nel senso che, se lo si suppone dotato di una memoria virtualmente infinita, un calcolatore di von Neumann è in grado di calcolare tutte le funzioni computabili secondo la Tesi di Church (per questo si dice che una macchina di von Neumann è un *calcolatore universale*). Per queste ragioni, la MTU costituisce un modello astratto degli attuali calcolatori digitali (elaborato prima della loro realizzazione fisica), e la teoria delle MT e della ricorsività costituisce, per così dire, la dottrina dei fondamenti dell'informatica teorica.

Per altre vie, la teoria delle MT ha avuto influenze rilevanti sulle riflessioni filosofiche sulla natura della mente. La Tesi di Church gode di uno statuto particolare nell'ambito degli enunciati matematici. "Algoritmo" e "funzione computabile in modo algoritmico" sono concetti intuitivi, non specificati in modo formale, per cui una dimostrazione rigorosa di equivalenza con il concetto di funzione calcolabile da una MT non è possibile. La Tesi di Church quindi non è un teorema, e neppure una congettura che potrebbe un giorno, in linea di principio, diventarlo. Si tratta di un enunciato che si fonda su un ampio spettro di evidenza matematica di tipo euristico⁵. Ai nostri fini, è interessante notare che gli approcci al problema della computabilità basati sull'elaborazione di macchine astratte, come appunto le MT, hanno condotto alcuni studiosi a considerare la Tesi di Church come una sorta di "legge empirica" piuttosto che come un enunciato a carattere logico-formale. Il logico Emil Post, il quale propose un concetto di macchina calcolatrice in parte analogo a quello sviluppato da Turing (Post 1936), sottolineava il suo disaccordo con chi tendeva ad identificare la Tesi di Church con un assioma o una mera definizione. Essa dovrebbe piuttosto essere considerata, afferma Post, una "ipotesi di lavoro", che, se opportunamente "corroborata", dovrebbe assumere il ruolo di una "legge naturale", una "fondamentale scoperta circa le limitazioni del potere matematizzante dell'*Homo sapiens*".

Su questa linea procedono alcuni sviluppi successivi del pensiero dello stesso Turing. Nel saggio "Macchine calcolatrici ed intelligenza" (Turing 1950) assistiamo ad una sorta di radicalizzazione del modo di intendere la Tesi di Church. Turing si dichiara fiducioso che un calcolatore universale possa giungere a simulare, nel volgere di pochi decenni, non soltanto il comportamento computazionale ed algoritmico di un essere umano, ma anche qualsiasi altra attività umana di tipo linguistico. Turing propone di riformulare la domanda "possono pensare le macchine?" nei termini del cosiddetto *gioco dell'imitazione*. Il gioco viene giocato da tre "attori": a) un essere umano, b) una macchina calcolatrice e c) un altro essere umano, l'interrogante. L'interrogante non può vedere a) e b), non sa chi dei due sia

⁵Che non è possibile analizzare in questa sede; si veda però (Kleene 1952).

l'uomo, e può comunicare con loro solo in maniera indiretta (ad esempio, supponendo di disporre della tecnologia attuale, attraverso un terminale video e una tastiera). L'interrogante deve sottoporre ad a) e a b) delle domande, in maniera tale da scoprire, nel più breve tempo possibile, quale dei due sia l'uomo e quale la macchina. a) si comporterà in modo da agevolare c), mentre b) dovrà rispondere in maniera tale da ingannare c) il più a lungo possibile. Invece di chiedersi se le macchine possono pensare, dice Turing, è più corretto chiedersi se una macchina possa ingannare un uomo nel gioco dell'imitazione, o, comunque, quanto a lungo possa resistergli. Questo "esperimento mentale" viene oggi abitualmente indicato col nome di *Test di Turing*.

Turing era eccessivamente ottimista circa le possibili prestazioni delle macchine calcolatrici: "credo che entro circa 50 anni sarà possibile programmare calcolatori [...] per far giocare loro il gioco dell'imitazione così bene che un esaminatore medio non avrà più del 70 per cento di probabilità di compiere l'identificazione esatta dopo cinque minuti di interrogazione. Credo che la domanda iniziale, 'possono pensare le macchine?', sia troppo priva di senso per meritare una discussione. Ciò nonostante credo che alla fine del secolo l'uso delle parole e l'opinione corrente saranno talmente mutate che chiunque potrà parlare di macchine pensanti senza aspettarsi di essere contraddetto" (in Somenzi e Cordeschi 1986, p. 166). Ci troviamo di fronte ad una sorta di versione "estremista", o "radicale", della Tesi di Church, che, grosso modo, potrebbe essere formulata come segue: *ogni attività linguistico-cognitiva è calcolabile da una macchina di Turing* (il che non vuol dire, ovviamente, che la nostra mente funziona *come* una macchina di Turing, ma che ogni attività mentale è simulabile da un dispositivo che abbia la stessa potenza computazionale di una macchina di Turing).

Questo testo viene spesso considerato uno degli atti di nascita dell'intelligenza artificiale. Lo stesso test di Turing è stato a lungo considerato come il criterio più adeguato per stabilire la validità empirica dei modelli computazionali della mente. In realtà è ancora preistoria, ancora pesantemente condizionata da assunzioni di tipo comportamentista. Da qui parte tuttavia una delle tradizioni dello studio computazionale della mente. La versione "estremista" della Tesi di Church è di fatto condivisa, in qualche forma, dalla maggior parte dei seguaci dell'approccio computazionale classico all'intelligenza artificiale e alla scienza cognitiva, seppure fondata su basi teoriche più articolate.

Sul versante filosofico, tali basi sono state fornite dalla *filosofia funzionalista della mente*. Negli anni successivi alla seconda guerra mondiale, le due tendenze allo studio filosofico della mente più diffuse nel mondo anglosassone erano il comportamentismo e il riduzionismo di tipo materialista (nella forma della teoria dell'identità degli stati mentali con stati del sistema nervoso centrale). Il funzionalismo nasce come reazione a entrambe queste posizioni. Prendiamo come punto di partenza il problema di stabilire che cosa significa dire che due soggetti si trovano nello stesso stato mentale, oppure che di essi è vero un certo predicato psicologico, come ad esempio *soffrire* o *credere che p* (dove *p* è un certo enunciato). Il *comportamentismo* (o *behaviorismo*) è una teoria filosofica della mente legata alle posizioni del neopositivismo logico. La risposta dei comportamentisti al problema degli stati mentali deriva direttamente dalla teoria neopositivistica della conoscenza e del significato: poiché le uniche espressioni linguistiche dotate di senso sono quelle riducibili a resoconti dell'esperienza empirica, il discorso psicologico non può fare riferimento a entità mentali interne inaccessibili dal punto di vista osservativo (stati mentali, rappresentazioni mentali, pensieri, e così via). Quindi, l'unico modo per conferire un senso a un linguaggio psicologico consiste nel fonderlo sui comportamenti esibiti in modo manifesto. Il comportamentismo presenta due varianti principali, il *comportamentismo logico*, sviluppato in ambito filosofico, e il *comportamentismo metodologico*, nato nell'ambito della ricerca psicologica empirica. Lo scopo dei *comportamentisti logici* (Ryle 1949; Hempel 1980) era quello di fornire dei criteri di significato accettabili dal punto di vista neoempirista per quelle espressioni del linguaggio ordinario che concernono fenomeni mentali: si trattava di individuare insiemi di condizioni necessarie e sufficienti affinché le espressioni linguistiche riferite a stati mentali potessero essere ridotte ad espressioni concernenti il comportamento. I *comportamentisti metodologici* (Skinner 1953) si proponevano invece di fornire una metodologia della psicologia in quanto scienza: il loro scopo era di fondare una psicologia scientifica in accordo con i dettami del neoempirismo, a prescindere dal fatto che ciò comportasse o meno un recupero e una giustificazione del linguaggio psicologico ordinario e del senso comune.

La teoria dell'*identità* degli stati mentali con stati del sistema nervoso centrale (ad es. Smart 1959; Armstrong 1981) è un tipo di riduzionismo materialista che storicamente si è contrapposto alle varie forme di comportamentismo. I seguaci della teoria dell'identità sostengono che è legittimo postulare l'esistenza di stati mentali e di oggetti mentali interni, e propongono di identificarli con stati cerebrali descrivibili a livello neurofisiologico. Ad ogni stato psicologico (ad esempio *credere che p*) dovrebbe corrispondere un predicato (diciamo C_p) definibile nel linguaggio della neurofisiologia. Dire che qualcuno, in un certo istante, crede *p* dovrebbe essere equivalente a dire che, in quell'istante, C_p è vero del suo cervello.

Sia comportamentismo che teoria materialista dell'identità presentano numerose difficoltà dal punto di vista filosofico. Un problema che i comportamentisti logici dovettero porsi da subito concerne il fatto che il comportamento tipico corrispondente ai vari termini psicologici emerge solo a determinate condizioni: ad esempio, qualcuno può provare dolore, o avere fame, e non potere, o non volere, esibire il comportamento corrispondente. D'altro canto, un comportamento può essere esibito fingendo. La soluzione proposta fu quella di ammettere la possibilità che le traduzioni dal linguaggio mentale al linguaggio comportamentale comprendessero dei condizionali con antecedenti di tipo controfattuale. Ad esempio, provare dolore significa che, *se* si verificano certe condizioni opportune (mancanza di vincoli e impedimenti di altro tipo, eccetera), *allora* viene esibito un certo comportamento. Ciò comporta tuttavia ulteriori problemi. In filosofia della scienza le proprietà che ammettono traduzioni di questo tipo vengono dette

proprietà disposizionali. A prescindere dalla difficoltà di individuare condizioni controfattuali opportune per la definizione comportamentista dei termini mentali, è noto che le definizioni di questo tipo non consentono una piena eliminabilità del *definiendum* (Hempel 1952, par. 6). Quindi, esse non consentono una riduzione completa del discorso mentale nei termini di quello comportamentale, come era nel progetto originario del comportamentismo logico.

Per quanto concerne la teoria materialista dell'identità, essa comporta ad esempio un forte "sciovinismo" antropocentrico: per avere stati mentali bisogna avere un sistema nervoso come quello dell'uomo. Esseri diversi (animali, macchine, angeli, extraterrestri) non potrebbero avere stati mentali, o comunque non potrebbero avere gli stessi stati mentali di un essere umano⁶.

I filosofi funzionalisti proposero di superare queste difficoltà ipotizzando che i criteri per l'identificazione degli stati mentali vadano individuati in base alle relazioni funzionali fra gli stati mentali stessi. Il funzionalismo riconosce la legittimità di postulare l'esistenza di stati mentali interni; tuttavia tali stati interni non vengono identificati con stati descrivibili a livello neurofisiologico, ma vengono definiti esclusivamente dal punto di vista funzionale, sulla base cioè delle loro relazioni reciproche e delle loro relazioni con gli *input* sensoriali e gli *output* comportamentali: due stati sono identici se sono identiche le loro relazioni funzionali con gli altri stati e con gli *input* e gli *output* del sistema, a prescindere da come essi sono realizzati fisicamente. Questo è esattamente il modo in cui sono definiti gli stati interni di una macchina calcolatrice astratta, ad esempio di una MT. Le macchine calcolatrici astratte, le MT in particolare, diventano quindi il modello secondo il quale viene pensata la mente da parte di alcuni filosofi funzionalisti. Rispetto alla teoria materialista dell'identità, il funzionalismo presenta il vantaggio che uno stesso stato mentale può essere "istanziato" da supporti fisici completamente diversi purché idonei a realizzare le relazioni funzionali rilevanti: in linea di principio una mente potrebbe essere fatta di formaggio svizzero, dice ad esempio Putnam. Facendo riferimento alla terminologia informatica, la definizione funzionalista degli stati mentali porta all'identificazione mente/*software* e sistema nervoso/*hardware*: come le proprietà formali che definiscono un programma prescindono dalle caratteristiche fisiche dei calcolatori che lo eseguono, così le proprietà degli stati mentali prescindono dalle caratteristiche fisiche ed anatomiche del sistema nervoso. La mente può essere vista come una macchina virtuale astratta implementata sull'*hardware* neuronale. Nasce quindi la metafora della mente come programma, o come insieme di programmi, che "girano" sul cervello.

Fra la fine degli anni cinquanta e gli anni sessanta (e soprattutto a partire dalla fine degli anni sessanta) gli sviluppi tecnologici dell'informatica fornirono gli strumenti perché le tesi del funzionalismo potessero trasformarsi, da pura speculazione filosofica, ad una ipotesi che guidasse la realizzazione effettiva di programmi per calcolatore che simulassero attività di tipo mentale, dando origine alla disciplina che va sotto il nome di *intelligenza artificiale*⁷. Il termine "intelligenza artificiale" si può far risalire all'estate del 1956, quando, al Dartmouth College di Hanover, nel New Hampshire, si tenne un congresso dal titolo *The Dartmouth Summer Research Project on Artificial Intelligence*, i cui organizzatori principali furono John McCarthy e Marvin Minsky⁸.

La nascita dell'IA comportò una "riunificazione" delle due tradizioni che si erano sviluppate a partire dalle MT, vale a dire quella di ambito informatico e quella relativa alla filosofia della mente. Oltre a queste, erano presenti anche altre componenti culturali. Importanti contributi furono forniti, ad esempio, dalla psicologia cognitivista, dalla linguistica chomskiana, dalla logica matematica.

Ci fu da sempre una certa ambiguità di intenti circa lo statuto e i fini della ricerca in IA, dovuta in parte proprio alla natura eterogenea delle discipline che vi sono confluite. Da un lato la ricerca in IA si è configurata come uno studio di tipo empirico sulle attività cognitive umane, quindi campo di ricerca di psicologi interessati alla elaborazione di modelli computazionali (per questa prospettiva si veda ad esempio Greco 1988). Per altri versi è stata posta l'enfasi sull'indagine di strumenti di rappresentazione, linguaggi e tecniche computazionali, che hanno portato ad identificare certe parti dell'IA con un ramo dell'informatica teorica. Vi è sempre stata infine una forte componente di tipo applicativo, che avvicinava l'IA a una disciplina di tipo ingegneristico.

In ogni caso, l'assunto universalmente condiviso da chi opera nell'ambito dell'IA classica è che un trattamento adeguato del comportamento intelligente presuppone la facoltà di manipolare rappresentazioni simboliche secondo regole di elaborazione di tipo formale. L'IA presuppone infatti una *teoria rappresentazionale della mente* secondo cui la mente costruisce rappresentazioni del mondo esterno, che svolgono un ruolo fondamentale nel pensiero e nel determinare il comportamento⁹. Questo è l'assunto fondamentale di quella che viene usualmente detta *psicologia del*

⁶ Per maggiori dettagli sulle critiche filosofiche al comportamentismo e alla teoria dell'identità rimandiamo a (Fodor 1980b) e ai saggi in (Putnam 1975).

⁷ Sui concetti e sui fondamenti teorici dell'intelligenza artificiale si veda ad esempio (Haugeland 1985). Una raccolta di contributi classici del settore è (Webber e Nilsson 1981).

⁸ Per questi avvenimenti e, in generale, per la storia delle origini dell'IA si può consultare (McCorduck 1979).

⁹ Alle teorie rappresentazionali della mente si contrappongono le teorie di tipo *eliminativista*, in base alle quali i concetti della psicologia di senso comune, primo fra tutti il concetto di rappresentazione mentale, sono concetti ingenui, prescientifici, che non corrispondono ad alcuna realtà di fatto, e che devono quindi essere eliminati da ogni discorso rigoroso, a carattere scientifico o filosofico, sulla mente. Una posizione filosofica di tipo eliminativista è ad esempio

sensu comune (*folk psychology*). Secondo la psicologia del senso comune gli esseri intelligenti esibiscono certi comportamenti perché hanno determinati desideri e determinate credenze: un politico conduce la propria campagna elettorale in un certo modo perché desidera essere eletto, e perché sa, o crede, che l'elettorato è particolarmente sensibile a certi problemi; una tigre si apposta vicino ad una sorgente perché desidera catturare una preda, e perché sa, o crede, che la preda verrà ad abbeverarsi. In questa prospettiva, gli stati mentali sono essenzialmente entità di tipo rappresentazionale: gli oggetti della credenza e del desiderio (come essere eletto, o che la preda venga ad abbeverarsi) sono rappresentazioni mentali di stati possibili o reali del mondo esterno. Pensare equivale ad elaborare tali rappresentazioni. L'IA, così come la scienza cognitiva e la maggior parte delle filosofie funzionaliste della mente, condivide questi assunti con la psicologia di senso comune.

Inoltre, in IA gli stati mentali vengono identificati con rappresentazioni di tipo simbolico. Uno dei "manifesti teorici" di questa maniera di concepire la mente e il pensiero è costituito dalla cosiddetta *ipotesi del sistema simbolico fisico* di Newell e Simon (1976). Un *sistema simbolico fisico* viene definito da Newell e Simon come un insieme di entità, i simboli appunto, che possono essere combinate in strutture simboliche complesse. Inoltre, il sistema dispone di processi che consentono di generare, trasformare o distruggere tali strutture. I simboli possono essere in una relazione di *designazione* con altri simboli, con processi computazionali o con oggetti esterni al sistema. L'aggettivo *fisico*, infine, significa che un sistema del genere deve obbedire alle leggi della fisica, e quindi deve essere realizzabile per mezzo di un dispositivo meccanico (anche se resta valida l'assunzione funzionalista per cui non importa quale specifico dispositivo fisico realizzi un dato sistema). Quindi "un sistema simbolico fisico è una macchina che produce nel tempo un insieme di strutture simboliche in evoluzione, e un tale sistema esiste in un mondo di oggetti che è più ampio di tali espressioni simboliche" (Newell e Simon 1976, p. 48 della trad. it.). L'ipotesi del sistema simbolico fisico viene formulata come un'ipotesi empirica sulla natura dell'intelligenza, la quale afferma che "un sistema simbolico fisico è in possesso degli strumenti necessari e sufficienti per l'azione intelligente generale" (ibid., p. 49), dove "con 'necessario' intendiamo affermare che ogni sistema che esibisce intelligenza generale mostrerà ad una successiva analisi di essere anche un sistema simbolico fisico. Per 'sufficiente' si intende che ogni sistema simbolico fisico di dimensioni sufficienti può essere ulteriormente organizzato per esibire intelligenza generale" (ibid.).

In generale, dunque, si può assumere che, in un sistema di simboli come viene concepito in questo ambito, sia possibile distinguere fra un insieme di simboli primitivi, o atomici, e un insieme di simboli complessi, ottenuti a partire da quelli atomici mediante opportune regole di composizione sintattica. Vi sono poi procedure e regole, che consentono di manipolare i simboli in modo da trarre le conseguenze volute. Inoltre, i simboli sono dotati di significato, in modo tale che il significato dei simboli complessi dipenda dal significato dei loro componenti¹⁰. Le procedure e le regole devono elaborare i simboli in accordo con tale attribuzione di significato. Il fatto di disporre di regole di questo tipo è ciò che rende il sistema automatizzabile: esso può essere elaborato in maniera formale da un dispositivo meccanico che segue le regole sintattiche, senza prendere in considerazione il significato dei simboli. L'analogia fra le strutture mentali così intese e strutture di tipo linguistico è evidente: in IA, come nelle discipline della mente ad essa collegate (filosofia funzionalista della mente, psicologia cognitiva, etc.) il linguaggio ha sempre avuto un ruolo privilegiato, sia come oggetto di studio, sia come metafora per la comprensione dei fenomeni mentali; si consideri ad esempio l'ipotesi di un funzionalista come Fodor, per cui i costrutti mentali sarebbero espressioni di un *linguaggio del pensiero* (Fodor 1975).

Dato l'assunto della natura rappresentazionale e simbolica della mente, è ovvio che il problema di *come dovessero essere rappresentate* le informazioni in un sistema intelligente assumesse un ruolo assolutamente centrale nella ricerca in IA. Il settore di indagine che va sotto il nome di *rappresentazione della conoscenza* ha come scopo l'individuazione e lo studio di linguaggi formali adatti a questo fine¹¹.

Gli obiettivi della rappresentazione della conoscenza possono essere espressi in maniera più precisa facendo riferimento alla cosiddetta *knowledge representation hypothesis* formulata dallo studioso di IA Brian Smith (Smith 1982). Secondo Smith, ogni "sistema intelligente" deve incorporare un insieme di strutture di tipo, in senso lato, linguistico, tali che:

quella di Churchland (1986). Anche nell'ambito delle scienze cognitive e dell'IA in senso lato sono state sviluppate posizioni eliminativiste. L'esempio più noto è costituito probabilmente dalle teorie del *planning reattivo* e della *nuova robotica*. Secondo Rodney Brooks, il più noto rappresentante di questa tendenza, l'essenza dell'intelligenza non risiede tanto nella capacità di disporre e di elaborare rappresentazioni mentali complesse, quanto piuttosto nella capacità di interagire in modo appropriato con la complessità del mondo esterno (si veda ad es. Brooks 1991). Si noti infine che una teoria rappresentazionale della mente non si accompagna necessariamente all'ipotesi simbolica sulla natura delle rappresentazioni mentali. Il *connessionismo* costituisce l'esempio più noto di un paradigma rappresentazionale ma non simbolico in scienza cognitiva (si veda Smolensky 1988).

¹⁰Assumendo che i simboli siano dotati di significato adottiamo qui quella che nel par. 12.1 chiameremo *accezione semantica* di "simbolo". In quel paragrafo vedremo come il termine "simbolo" venga talvolta utilizzato in un'accezione più debole, di tipo esclusivamente sintattico.

¹¹(Ringland e Duce 1988) è un'introduzione alla rappresentazione della conoscenza; (Brachman e Levesque 1985) è una antologia di articoli di rappresentazione della conoscenza non molto recenti ma in molti casi ormai classici o storicamente importanti; un'altra raccolta di saggi che può risultare utile è (Cercone e McCalla 1987).

1. queste strutture, viste da un osservatore esterno al sistema, possono essere interpretate come la rappresentazione della conoscenza di cui il sistema dispone;
 2. indipendentemente da tale "attribuzione semantica" data dall'esterno, tali strutture devono essere manipolabili formalmente, in modo da poter giocare un ruolo causale nel determinare il comportamento del sistema.
- H. Levesque e R. Brachman (1985), altri due studiosi del settore, commentano l'ipotesi di Smith come segue:

[...] ci sono due proprietà principali che devono essere soddisfatte dalle strutture di un sistema basato sulla conoscenza.

Innanzitutto, deve essere possibile interpretarle come *proposizioni* che rappresentano la conoscenza globale del sistema. Altrimenti, potrebbe trattarsi di una rappresentazione non di *conoscenza*, ma di qualcosa di molto diverso, come numeri o circuiti. È implicito in questo vincolo che tali strutture debbano essere rappresentazioni in un linguaggio che abbia una *teoria della verità*. Per ognuna di esse, dovremmo essere in grado di dire come deve essere il mondo perché essa sia vera. Tali strutture non devono *assomigliare* a frasi (non c'è nessun vincolo sintattico su di esse, a parte forse la finitezza); tuttavia dobbiamo essere in grado di comprenderle come se lo fossero.

Un secondo requisito dell'ipotesi è forse più ovvio. Le strutture simboliche in una certa base di conoscenza devono giocare un ruolo *causale* nel comportamento del sistema, in contrapposizione, ad esempio, ai commenti in un linguaggio di programmazione. Inoltre, l'influenza che hanno sul comportamento del sistema deve concordare con la nostra comprensione di esse in quanto proposizioni che rappresentano conoscenza. Questo non vuol dire che il sistema debba essere cosciente in qualche modo misterioso dell'interpretazione delle sue strutture e della loro connessione con il mondo. Ma perché abbia senso considerarlo "basato sulla conoscenza" dobbiamo essere in grado di interpretare il suo comportamento come se comprendesse quelle proposizioni, allo stesso modo in cui interpretiamo il comportamento di un programma di tipo numerico come se esso fosse in grado di comprendere la relazione fra strutture di bit e quantità numeriche astratte. [...] In altre parole, il fine di un sistema di rappresentazione della conoscenza è di individuare *strutture simboliche e meccanismi di inferenza* appropriati sia per rispondere a domande che per acquisire nuove informazioni, in accordo con la teoria della verità del linguaggio di rappresentazione sottostante. (pp. 44-47)

Resta da stabilire in che modo tali strutture di rappresentazione debbano essere caratterizzate in maniera più precisa. I sistemi formali elaborati dai logici godono esattamente di tali caratteristiche. È quindi naturale che si sia pensato di utilizzarli a questo scopo. Tuttavia, tale proposta ha incontrato un consenso tutt'altro che unanime: il problema dell'impiego della logica in IA ha scatenato una lunga serie di dispute a carattere teorico e fondazionale.

3.2 Logicisti e Anti-logicisti

Quando si introducono tematiche di carattere logico nell'ambito dell'IA si entra in un terreno minato; in passato il ruolo degli strumenti logici è stato oggetto di innumerevoli discussioni, e ancora oggi si possono registrare opinioni non convergenti per quanto riguarda l'opportunità del loro impiego. In anni più recenti i termini della discussione si sono fatti più distesi e pacati. Tuttavia il problema del rapporto fra logica e rappresentazione della conoscenza resta uno dei problemi centrali dell'IA¹².

Già fin dagli anni cinquanta, John McCarthy, uno dei più autorevoli sostenitori dell'approccio logico all'IA, aveva profetizzato che la logica avrebbe dovuto svolgere nei confronti della *computer science*, e, in particolare, dell'IA, la funzione che l'analisi infinitesimale ha svolto (e svolge tuttora) nei confronti della fisica. Queste posizioni non suscitarono tuttavia il consenso dell'intera comunità dei ricercatori. Si formarono due differenti punti di vista, e la ricerca si sviluppò lungo linee divergenti. McCarthy e Marvin Minsky possono essere considerati rispettivamente i sostenitori più rappresentativi delle due opposte tendenze. McCarthy ritiene che lo scopo dell'IA sia progettare programmi per calcolatore che funzionino in accordo coi dettami della logica matematica, a prescindere dal fatto che questo sia o meno il modo in cui ragionano gli esseri umani. Minsky pensa che l'approccio migliore sia fare sì che i

¹²Recentemente, un numero monografico della rivista *Sistemi Intelligenti* (1991) è stato dedicato al problema del ruolo della logica in IA, e può costituire un utile punto di riferimento. Si veda in particolare (Lolli 1991). Un manuale di IA che segue l'impostazione logica è Genesereth e Nilsson (1987). I vari manuali elaborati da Nils Nilsson nel corso degli anni sono in un certo senso sintomatici dell'evoluzione dell'IA simbolica nella direzione logicista. Nilsson (1971) poneva l'enfasi sui metodi di ricerca nello spazio degli stati, riservando una trattazione relativamente esigua alla logica. In (Nilsson 1980) le proporzioni sono invertite, e viene dato ampio spazio ad aspetti computazionali del calcolo dei predicati, come ad esempio le diverse strategie di risoluzione. Genesereth e Nilsson (1986) infine estendono la trattazione della logica introducendo numerose tematiche di rappresentazione della conoscenza e di ragionamento del senso comune, quali il ragionamento non monotono, l'induzione, il ragionamento incerto, le credenze, e così via.

calcolatori imitano il modo in cui funziona la mente umana, che, ritiene Minsky, è certamente diverso dal modo in cui funziona la logica matematica (si veda Kolata 1982).

Questa contrapposizione ha avuto luogo in praticamente tutti i settori di ricerca dell'IA. Il settore della dimostrazione automatica di teoremi è uno dei più "antichi" dell'IA e quello in cui gli informatici si sono trovati ad operare in un terreno già ampiamente esplorato dai logici: l'elaborazione di calcoli corretti e completi in cui le formule logicamente valide sono derivabili con procedimenti meccanici aveva mostrato come un'attività di tipo algoritmico poteva surrogare il pensiero umano. Tuttavia, i limiti posti dai teoremi di indecidibilità di Turing-Church e dai problemi computazionali connessi con la lunghezza delle dimostrazioni e con l'esplosione combinatoria hanno lanciato una nuova sfida ai ricercatori. Anche qui, per risolvere tali problemi, si è manifestata un'opposizione tra due diverse tendenze: la prima si è rivolta al perfezionamento dei calcoli logici esistenti, l'altra, pur volendo impiegare il calcolatore per simulare l'attività dimostrativa, si oppone ai calcoli logici già precedentemente elaborati per fini diversi da quelli implementativi, rivolgendosi alla ricerca di strategie maggiormente *human oriented*, ispirate da considerazioni di ordine psicologico (Lolli 1988). Per esempio, la *Geometry Proving Machine* di Gelearneter e Gilmore (Gelearneter 1963; Gilmore 1970) è basata sull'idea di Minsky di utilizzare la costruzione geometrica di figure come guida euristica nella dimostrazione di teoremi geometrici. Al di sotto di questa opposizione si ripropone la distinzione di fondo delle due diverse ottiche, quella della *performance mode*, orientata più verso i risultati che ai metodi impiegati, e quella della *simulation mode*, che pone al centro dei suoi interessi la simulazione dei processi mentali dell'uomo.

Nell'ambito specifico della rappresentazione della conoscenza i "logicisti" proponevano di utilizzare gli strumenti della logica matematica per studiare e sviluppare i linguaggi di rappresentazione della conoscenza, mentre il versante "anti-logico" può essere identificato, in senso lato, con i sostenitori di sistemi di rappresentazione alternativi quali i *frame* e le reti semantiche. Il punto di partenza principale degli "anti-logici" è consistito nel mettere in evidenza le differenze fra logica matematica e ragionamento di senso comune: la logica è stata sviluppata per fare ciò che il senso comune non è in grado di fare in quanto vago e contraddittorio, ma se vogliamo che un calcolatore "pensi" come un essere umano, dobbiamo rivolgerci a quelle caratteristiche tipiche del ragionamento di senso comune che non sono state prese in considerazione, o che sono state deliberatamente rifiutate, dalla logica matematica¹³.

Oltre alle due posizioni estreme, ne esistono altre intermedie e più sfumate. Ad esempio, nel settore della rappresentazione della conoscenza, secondo Israel e Brachman (1981) può essere individuato il seguente spettro di possibili posizioni riguardo al problema dell'uso delle reti semantiche rispetto alla logica:

1. *logicists*: i logici "puri", che caldeggiavano l'uso dei calcoli logici tradizionali direttamente come linguaggi di rappresentazione della conoscenza;
2. *net-logicists*: considerano le reti semantiche come semplici varianti notazionali della logica del primo ordine, in grado di offrire alcuni vantaggi dal punto di vista implementativo;
3. *extended-net-logicists*: considerano le reti semantiche come una variante notazionale dei calcoli logici, con tuttavia qualche vantaggio dal punto di vista espressivo;
4. *net-workers*: sostengono che le reti semantiche non sono riconducibili alla logica. Possono essere ulteriormente suddivisi in due gruppi:
 - *psycho-net-workers*: sostengono che le reti semantiche devono essere empiricamente adeguate da un punto di vista psicologico, o semplicemente influenzate da considerazioni di carattere psicologico.
 - *not-psycho-net-workers*: sostengono che le reti semantiche sono formalismi universali di rappresentazione, che non devono essere soggetti a vincoli di carattere psicologico.

Ciascuna di queste posizioni è stata sostenuta e discussa nella letteratura. Tuttavia, nel resto di questo capitolo, prenderemo in considerazione come paradigmatiche solo le due posizioni più radicali. Nel seguito si esporranno le principali obiezioni all'uso della logica in IA, e quindi alcune delle risposte dei logicisti a tali obiezioni.

Veniamo dunque agli argomenti che motivano le posizioni degli "anti-logici". I teorici della IA hanno iniziato negli anni sessanta a porsi il problema di progettare sistemi simbolici mediante i quali affrontare il problema del ragionamento di senso comune, andando alla ricerca di a) un linguaggio formalizzato con il quale rappresentare la

¹³ Nelle righe precedenti il termine *logicists* è stato messo fra virgolette per sottolineare il fatto che qui non viene utilizzato con il significato che ha usualmente in filosofia della matematica, dove indica la corrente che ha cercato di dare un fondamento alla matematica riducendola alla logica. Si noti tuttavia che Thomason (1991) ravvisa delle somiglianze fra il programma logicista in IA ed altre forme di logicismo - non soltanto il logicismo in filosofia della matematica, ma anche quello che Thomason chiama "logicismo linguistico", ossia appunto il paradigma model teoretico in filosofia del linguaggio.

conoscenza del soggetto da modellare, b) una rappresentazione dettagliata della conoscenza di base di un tale soggetto, e c) un meccanismo mediante il quale si possano inferire in modo automatico le conoscenze implicite nella conoscenza di base. Le obiezioni all'uso della logica in questa impresa possono essere molteplici. Il problema è che, storicamente, le finalità con cui sono stati sviluppati i calcoli logici sono radicalmente diverse. Innanzi tutto, come abbiamo visto, la logica moderna, da Frege in poi, è sempre stata guidata da un forte assunto anti-psicologista. Lo scopo del logico non è quello di modellare il modo in cui un essere umano pensa, ma è quello di costruire un modello astratto e formale dei procedimenti inferenziali. La logica non dice come viene effettuato un ragionamento, ma come un'inferenza già data può essere giustificata come valida. La logica si è posta quindi, in origine, un intento di tipo normativo: l'elaborazione di un linguaggio perfetto ideale che non soffrisse di tutte le lacune e le ambiguità delle lingue naturali, e di regole di inferenza definite su tale linguaggio. Inoltre il punto di riferimento che era stato assunto per la realizzazione di questo programma era il tipo di ragionamento in uso nelle scienze matematiche: la logica, almeno inizialmente, è nata come la formalizzazione del ragionamento matematico. L'interesse verso questo ordine di problemi è una delle ragioni che hanno indotto molti teorici dell'IA a sottolineare la differenza di obiettivi tra le ricerche svolte in logica matematica e quelle svolte nel settore della rappresentazione della conoscenza, essendo queste ultime orientate a elaborare formalismi che intendono riprodurre il funzionamento della mente umana. Dice Minsky (1982): "Molti ricercatori dell'IA hanno continuato a perseguire l'uso della logica per la risoluzione di problemi. Ritengo che questo non abbia dato buoni frutti. Il ragionamento logico è più adeguato per spiegare o confermare il *risultato* del pensiero. Io credo che noi usiamo la logica, piuttosto che per risolvere problemi, per spiegarne le soluzioni ad altre persone e - più ancora - a noi stessi".

Ragionamento e dimostrazione sono due concetti ben distinti. Nei confronti dei sistemi inferenziali da loro elaborati, i logici sono interessati soprattutto a proprietà metateoriche come la completezza e la correttezza. Secondo Minsky l'applicazione di regole di inferenza corrette e complete non è di importanza fondamentale nel *commonsense reasoning*. Il ragionamento di senso comune può condurci da premesse vere a conclusioni false, oppure consentirci inferenze corrette utilizzando regole tutt'altro che valide.

Una delle esposizioni più articolate delle "debolezze" insite nell'impiego dei formalismi logici risale al 1975, ed è l'appendice all'articolo di Minsky "A framework for representing knowledge". Le conclusioni sono le seguenti:

1. Il ragionamento "logico" non è abbastanza flessibile da servire come base per il pensiero: io preferisco considerarlo come un gruppo di metodi euristici, che diventano efficaci solo quando vengono applicati a piani schematici rigidamente semplificati. Altrimenti, la coerenza che la logica richiede non è di solito disponibile - *e probabilmente neppure desiderabile* - perché i sistemi coerenti probabilmente sono troppo deboli.
2. Io dubito della possibilità di rappresentare effettivamente la conoscenza ordinaria nella forma di molte brevi proposizioni, vere indipendentemente l'una dall'altra.
3. La strategia di separare completamente la conoscenza specifica dalle regole generali di inferenza è troppo radicale. Abbiamo bisogno di una maniera più diretta per collegare frammenti di conoscenza alle intuizioni su *come* usarla.
4. E' stato creduto a lungo che fosse essenziale rendere tutta la conoscenza accessibile alla deduzione nella forma di asserzioni dichiarative, ma questo sarebbe molto meno urgente nella misura in cui imparassimo a manipolare descrizioni strutturali e procedurali. (Minsky 1975, p. 262 di Brachman e Levesque 1985)

Analoghe critiche all'approccio logico sono mosse da Schank e Rieger (1974), che pongono l'accento sulla differenza fra il concetto di *inferenza* intesa come processo psicologico, e il concetto logico-formale di *deduzione*. Ad esempio, le inferenze spesso non sono deduzioni *valide* dal punto di vista logico. Un frammento di informazione ottenuto mediante inferenza non è mai da intendersi come vero in modo certo, ma solo plausibile ad un certo grado, e sempre in linea di principio rivedibile. Inoltre, l'inferenza è sempre guidata dai legami associativi fra i concetti presenti nella memoria.

E' fuori di dubbio che queste osservazioni sono pertinenti e centrate, e che i problemi sollevati riguardano, almeno in parte, anche gli approcci, come quello di McCarthy, in cui non si ritiene rilevante l'adeguatezza psicologica dei modelli elaborati. Esaminiamo quindi più in dettaglio alcuni punti specifici.

• *Prototipi contro concetti definiti*

Si tratta del corrispettivo, in rappresentazione della conoscenza, del problema posto in filosofia del linguaggio dai postulati di significato tradizionali. In logica i concetti devono essere definiti in maniera netta, in termini di condizioni necessarie e/o sufficienti, mentre il ragionamento di senso comune adopera concetti vaghi, spesso caratterizzati mediante prototipi: prendendo in considerazione un concetto (o un insieme) noi spesso non diamo definizioni esatte o complete delle proprietà degli individui che cadono sotto quel concetto (o dei membri dell'insieme), ma concentriamo la nostra attenzione su di un membro tipico, rappresentativo dell'insieme (un *prototipo*, appunto). Gli insiemi *basati su esempi* sembrano essere una caratteristica tipica del pensiero umano, anche dal punto di vista

psicologico. Questo è connesso con il problema dei *default value*, valori che vengono attribuiti alle istanze di un concetto in assenza di altre informazioni esplicite. Essi rappresentano le proprietà del prototipo e delle istanze tipiche di un concetto; è tuttavia possibile accettare istanze del concetto che hanno proprietà differenti. Ad esempio, gli elefanti tipici sono grigi ed hanno quattro zampe: queste sono proprietà del prototipo di elefante, assunte come vere per *default* per tutti gli elefanti "normali". Vi possono essere tuttavia istanze del concetto "elefante" che violano tali attributi (ad es. elefanti rosa o elefanti con tre zampe). Questo aspetto rimanda al problema del trattamento delle eccezioni e alla non-monotonicità del *commonsense reasoning*. Dice Minsky:

Si consideri un fatto come "gli uccelli possono volare". Se pensate che il *commonsense reasoning* sia simile al ragionamento logico, allora dovete credere che vi siano principi generali che affermano "se Joe è un uccello e gli uccelli possono volare, allora Joe può volare". Ma supponete che Joe sia uno struzzo o un pinguino. Bene, possiamo assiomaticamente dire che se Joe è un uccello, e Joe non è uno struzzo o un pinguino, allora Joe può volare. Ma se Joe è morto? O se ha i piedi nel calcestruzzo? Il problema con la logica è che una volta che voi avete dedotto qualcosa, non potete più sbarazzarvene. Dove voglio arrivare è che c'è un problema con le eccezioni.

Tradizionalmente in logica vale la proprietà di monotonicità: se una proposizione è conseguenza logica di un insieme di altre, è ancora, a maggior ragione, conseguenza logica di un soprainsieme di esse (cfr. cap. 6). All'opposto, il *commonsense reasoning* è tipicamente non monotono. Noi dobbiamo continuamente trarre conclusioni sulla base di conoscenza incompleta, ed essere disposti a ritrarre tali conclusioni se esse non sono più in accordo con le nuove conoscenze disponibili. Impariamo dal mondo, e dobbiamo rivedere assunzioni e credenze in precedenza accettate.

- *Inferenze "giuste" contro inferenze "sbagliate"*

Si deve considerare il fatto che gli esseri umani eseguono inferenze che sono spesso sistematicamente "sbagliate" dal punto di vista logico¹⁴. Si ricordino a questo proposito le osservazioni di Schank riportate più sopra. Inoltre, spesso accade che le credenze di un individuo siano contraddittorie, senza che questo abbia gli effetti disastrosi che una contraddizione comporta in logica (in un sistema logico classico in cui sia presente una contraddizione è deducibile qualsiasi formula). In molti casi tali "errori" possono non costituire un limite per un soggetto razionale finito. Per un soggetto limitato (sia dal punto di vista della conoscenza disponibile che delle capacità inferenziali) trarre certe inferenze in modo logicamente "corretto" potrebbe essere impossibile o troppo dispendioso in termini di tempo e di risorse computazionali. Ai fini di un comportamento globale efficiente ed adeguato, può quindi essere più vantaggioso e più "razionale" saltare alle conclusioni anche se in modo azzardato. Anche il problema del ragionamento non monotono nasce, in un certo senso, da questo ordine di problemi.

- *Struttura della memoria*

Non è plausibile che le informazioni nella memoria umana siano rappresentate in forma analoga ad una serie di assiomi logici. I dati disponibili fanno supporre che esse siano strutturate in maniera più complessa, organizzate in base al loro contenuto, in modo da renderne più agevole l'accesso per lo svolgimento dei vari compiti cognitivi (come il ragionamento, la comprensione del linguaggio, la percezione). Secondo Minsky, le informazioni nella nostra memoria sono rappresentate sotto forma di blocchi più estesi e più organizzati di singole formule isolate. Un formalismo adeguato di rappresentazione della conoscenza dovrebbe consentire di definire strutture analoghe. Inoltre per Minsky si dovrebbe superare una distinzione troppo netta fra conoscenza espressa in forma dichiarativa e procedure di inferenza valide in generale: ad ogni blocco di informazioni dovrebbero essere associati procedimenti inferenziali specifici.

Nel modellare numerose attività cognitive sono rilevanti i legami associativi di "vicinanza semantica" fra i concetti presenti nella memoria, legami che non sono riconducibili a nessi di tipo puramente logico. Ad esempio, il concetto *cometa* è più "vicino" al concetto *Natale* di quanto non lo sia il concetto *forchetta*, anche se questo non rispecchia nessuna relazione logica specifica. I costrutti in una base di conoscenza dovrebbero rispecchiare anche questa struttura associativa. In una base di conoscenza tutte le informazioni connesse ad una certa entità concettuale dovrebbero essere accessibili da un luogo unico. Strutture di rappresentazione come i *frame* e le reti semantiche avrebbero dovuto fornire una risposta a questo genere di problemi.

I "logicisti" non misconoscevano la rilevanza dei problemi posti dai loro avversari. Lo testimonia l'enfasi che essi hanno posto da sempre sul ragionamento non monotono. Ad esempio, in ambito logicista è stato individuato uno dei problemi tipici posti dalla non monotonicità del ragionamento di senso comune, il cosiddetto *frame problem* (McCarthy

¹⁴In questo settore molte ricerche sono state condotte dagli psicologi; si vedano ad esempio i lavori di Wason e di Johnson-Laird (Wason e Johnson-Laird 1972; Johnson-Laird 1983).

e Hayes 1969), sul quale torneremo in seguito¹⁵. I logicisti si limitavano a negare l'adeguatezza delle soluzioni proposte dai loro avversari, ritenendo troppo frettolosa da parte loro la liquidazione degli strumenti della logica. I problemi posti dagli oppositori della logica erano reali. Tuttavia i loro argomenti, almeno nei termini in cui erano stati originariamente formulati, non offrivano alcuna alternativa rigorosa possibile, se non in maniera informale ed alquanto schematica. Qui di seguito esporremo alcune delle obiezioni dei logicisti a Minsky e agli anti-logici. In primo luogo, sebbene l'introduzione dei sistemi logico-formali e dello studio delle loro proprietà metateoriche abbia avuto storicamente origine nell'ambito delle indagini sui fondamenti della matematica, abbiamo visto che la formalizzazione logico-matematica è stata progressivamente estesa a comprendere modalità, eventi, azioni, contesti spazio-temporali, ecc., rivelando una notevole flessibilità nel rendere esplicita la forma logica di classi sempre più ampie di enunciati del linguaggio comune. Storicamente, la logica matematica è stata sviluppata con lo scopo di determinare un linguaggio artificiale senza la vaghezza e le ambiguità del linguaggio naturale al fine di perseguire il massimo rigore dimostrativo. Tuttavia, in tempi più recenti, soprattutto nell'ambito delle logiche filosofiche e della filosofia del linguaggio di tradizione analitica, molti sforzi sono stati indirizzati verso la formalizzazione dello stesso linguaggio naturale e verso lo studio della sua semantica. Si è così aperta la possibilità di concepire i linguaggi formali non in opposizione ai linguaggi naturali, ma piuttosto come dinamiche approssimazioni alla ricchezza espressiva dei secondi e, in quanto tali, suscettibili di articolazioni e adattamenti in funzione delle più diverse situazioni che si presentano in ragionamenti condotti nel linguaggio naturale. Vediamo allora alcuni degli argomenti che gli studiosi di impostazione logica hanno opposto alle critiche degli "anti-logici".

- *Necessità di una semantica*

Uno degli argomenti centrali dei sostenitori della logica è incentrato sulla necessità di un'adeguata semantica per i formalismi di rappresentazione della conoscenza. P. Hayes (1974) mise in luce il fatto che molti nuovi formalismi erano stati sviluppati senza una semantica, eliminando in questo modo ogni possibilità di seri confronti. Questo punto è stato ribadito da D. McDermott (1978), che ha sottolineato come, senza la possibilità di assicurare la correttezza dei costrutti utilizzati per mezzo di un'adeguata semantica, programmi lunghi e complessi rischiano di diventare una massa inestricabile di segni senza significato. McDermott fa l'esempio del sistema AMORD (deKleer et al. 1977). Generalmente, il sistema utilizza una versione della regola di risoluzione, che è ben studiata e ben compresa dal punto di vista logico. Tuttavia, ci sono casi nei quali le primitive del sistema non hanno una precisa interpretazione logica: in tali casi, la sola garanzia che il sistema sia usato correttamente sono l'intuizione e il buon senso dell'utente. Il fatto che un sistema di rappresentazione sia implementato da un programma scritto in un linguaggio come il LISP non garantisce nulla sul suo significato e sulla sua correttezza. A proposito del sistema di rappresentazione a *frame* KRL di Bobrow e Winograd (1977); McDermott (1978) afferma: "Un sistema come KRL, che consiste di uno splendido edificio di notazione senza denotazione, è un castello in aria". Ad esempio non è possibile stabilire quando in una base di conoscenza KRL c'è una contraddizione. Lo stesso vale per la maggior parte dei sistemi a rete semantica. Disporre di una semantica, dice McDermott, è cruciale anche per il problema dell'accrescimento di un sistema. Innanzi tutto si deve poter stabilire che cosa è esattamente in grado di fare il sistema esistente, e in che cosa esso è incompleto. Inoltre, si deve poter stabilire che l'aggiunta dei nuovi elementi non distrugga la correttezza di quello che già esiste.

Anche Moore (1982) ribadisce l'importanza di un'adeguata semantica: "Qualunque cosa un formalismo possa essere, almeno una parte delle sue espressioni deve avere una *semantica referenziale*, se il formalismo deve costituire la rappresentazione di qualche *conoscenza*. Cioè, deve esserci qualche tipo di corrispondenza fra un'espressione e il mondo, tale che abbia senso chiedersi se il mondo è realmente come quell'espressione lo descrive"(p. 428). Secondo Moore la logica matematica è la disciplina che si occupa esattamente delle relazioni fra espressioni simboliche e mondo. Nel momento in cui ci si pone il problema dell'analisi della semantica referenziale di un sistema di rappresentazione, si sta facendo logica: gli unici strumenti disponibili per questo tipo di analisi semantica sono quelli elaborati in logica formale.

- *Livello dell'espressione e livello dell'implementazione*

Nel passo seguente di William Woods viene posta una distinzione che, sotto varie forme e denominazioni, ha svolto un ruolo fondamentale nelle risposte di parte logica agli anti logici:

Due aspetti del problema della rappresentazione della conoscenza devono essere considerati. Il primo, che io chiamo *adeguatezza espressiva*, ha a che fare con il potere espressivo della rappresentazione - cioè con quello che può essere detto. [...] Il secondo aspetto, che io chiamo *efficacia notazionale*, riguarda la forma effettiva e le strutture della rappresentazione, e l'impatto di queste strutture sulle

¹⁵In breve, il *frame problem* ha a che fare con il ragionamento su tempo, azioni ed eventi. Quando si cerca di inferire le conseguenze di una data azione, si assume implicitamente che valgano certe condizioni "tipiche". Tuttavia, tali condizioni possono risultare false, per cui le conclusioni raggiunte devono essere ritirate.

operazioni del sistema. L'efficacia notazionale, a sua volta, si distingue in componenti quali l'efficienza computazionale rispetto a diversi tipi di inferenza, la sinteticità della rappresentazione, e la facilità di modificazione. E' importante distinguere adeguatezza espressiva da efficacia notazionale, poiché il fatto di non chiarire quale dei due problemi fosse affrontato ha esacerbato numerose dispute in questo campo (Woods 1983).

Alcuni autori hanno insistito sulla differenza fra *livello dell'espressione* (che concerne la progettazione di un linguaggio formale per esprimere conoscenze) e il *livello dell'implementazione* (che ha a che fare con i dispositivi computazionali che realizzano su un calcolatore tale linguaggio di rappresentazione). Il ruolo della logica si colloca al livello espressivo, e non a quello implementativo. Hayes (1977) sostiene che la logica non è uno stile di programmazione, e non comporta alcuna scelta a livello degli algoritmi implementativi:

La logica è una collezione di idee su come esprimere certi tipi di conoscenza a proposito di certe situazioni del mondo. La metateoria della logica è una collezione di strumenti matematici per analizzare questa classe di linguaggi di rappresentazione. Quello che questi strumenti analizzano non è il comportamento di un interprete, o la struttura di un processo in qualche sistema implementato, quanto piuttosto il significato estensionale delle espressioni di un linguaggio, qualora si assuma che esse facciano riferimento ad una realtà esterna. Questi due temi distinti - il significato di un linguaggio e il comportamento di un suo interprete - sono correlati in vari modi. Essi vengono in contatto soprattutto con la nozione di inferenza. Il significato logico *giustifica* certe inferenze. Un sistema implementato *esegue* delle inferenze: qualcuno dei suoi processi consiste nel fare delle inferenze. Ma sulla stessa nozione di inferenza e sullo stesso linguaggio di rappresentazione possono essere basati due sistemi diversi. La *struttura inferenziale* del linguaggio usato dal sistema non dipende dalla *struttura di elaborazione*. In particolare, un sistema può avere una struttura inferenziale di tipo logico - può effettuare inferenze deduttivamente valide - senza essere un classico dimostratore di teoremi che "macina assieme" liste di clausole (p. 559).

Analoga è la distinzione di McCarthy (McCarthy 1977; McCarthy e Hayes 1969) fra i problemi di tipo *epistemologico* e i problemi di tipo *euristico* dell'IA. I problemi epistemologici riguardano il tipo di conoscenza da rappresentare, la semantica e il potere espressivo dei linguaggi di rappresentazione, la correttezza delle regole di inferenza adottate. I problemi euristici riguardano invece la scelta delle strutture dati che implementano il linguaggio di rappresentazione, degli algoritmi che implementano le regole di inferenza, le strategie per esplorare gli spazi di ricerca, e così via. Secondo i sostenitori dell'approccio logico, molte delle obiezioni poste dai detrattori della logica non sono rilevanti a livello epistemologico, poiché riguardano esclusivamente gli aspetti euristici. Ad esempio, il fatto di raggruppare la conoscenza in unità complesse e strutturate come i *frame*, i vantaggi che offrono *frame* e reti semantiche nella consultazione di una base di conoscenza, o il problema di rappresentare la vicinanza semantica fra concetti sono riconducibili al problema euristico di come rendere facilmente accessibile la conoscenza rappresentata agli algoritmi che eseguono le inferenze, e non sono incompatibili con l'impiego della logica al livello epistemologico.

- *Logiche per il senso comune*

I logici concordano che il ragionamento umano è diverso dalla deduzione logica. Ad esempio, essi sono consapevoli che la storia della logica è dominata dalla ricerca di tipi di inferenza che preservino la verità, mentre spesso il ragionamento di senso comune non rispetta questo vincolo. Tuttavia si possono usare gli strumenti della logica per costruire modelli, più o meno astratti, del ragionamento di senso comune. Molti teorici dell'IA pensano che non ci sia nulla da opporre all'uso di linguaggi logici standard per affrontare questi problemi. E' anche possibile suggerire l'uso di sistemi non standard, come, ad esempio, le logiche paraconsistenti e le logiche condizionali. Moore (1982) sottolinea come il calcolo dei predicati sia molto flessibile, e ponga pochissime pregiudiziali "ontologiche" sulla struttura della realtà da rappresentare. Questo lo rende utilizzabile per formalizzare numerosi aspetti del ragionamento di senso comune:

Forse la caratteristica principale delle logiche del primo ordine è quella di descrivere il mondo in termini di oggetti e delle loro proprietà e relazioni. Io dubito che chiunque in IA possa trovare da ridire su ciò, poiché virtualmente tutti i formalismi di rappresentazione dell'IA fanno uso di questi concetti. Si potrebbe argomentare che, come nozioni primitive, serva qualcosa di più dei soli oggetti, proprietà e relazioni. Ma bisogna tenere presente che la logica del primo ordine non pone alcun limite su cosa debba essere considerato come oggetto. Non solo oggetti fisici, ma anche tempo, eventi, tipi, organizzazioni, mondi ed enunciati possono essere trattati come individui dal punto di vista logico. Inoltre, anche se decidiamo che ci servono caratteristiche "non-standard", quali ad esempio operatori intensionali o di ordine superiore, possiamo ancora incorporarli in un inquadramento logico. [...] La ragione principale per

cui considero le caratteristiche della logica del primo ordine come essenziali per ogni formalismo universale di rappresentazione è che esse si possono applicare alla conoscenza di *qualunque* dominio (p. 430).

Abbiamo già visto come, soprattutto in ambito filosofico, siano stati elaborati calcoli logici che modellano tipi di inferenza più vicini al senso comune che non al ragionamento matematico (è il caso delle logiche modali aleatiche, e di quelle deontiche, temporali, etc.). Vi sono poi tipi di problemi che non erano mai stati affrontati dai logici, ma che non escludono a priori un trattamento di tipo logico. Un esempio è costituito dal ragionamento non monotono. Tutte le logiche "tradizionali" sono monotone, sia la logica classica che le logiche non classiche (come, ad esempio, quella intuizionista). Tuttavia, in ambito IA, oggi sono in corso molte ricerche per sviluppare logiche non monotone (si veda oltre, il capitolo 6). In generale, la maggior parte della ricerca logica in IA è incentrata sul problema di elaborare modelli formali delle attività inferenziali di soggetti razionali finiti, non idealizzati, e limitati da diversi punti di vista (conoscenza, tempo e risorse di calcolo disponibili, etc.), problema di cui il ragionamento non monotono è solo un caso particolare. Ad esempio viene studiata la possibilità di individuare sistemi logici "deboli" adeguati a modellare le prestazioni di soggetti con capacità computazionali limitate (cfr. Levesque 1988). Su questi punti torneremo comunque in seguito.

Se si richiama la *knowledge representation hypothesis* di Brian Smith si ravvisano molte analogie fra la ricerca in rappresentazione della conoscenza e il lavoro dei logici. Si riveda ad esempio il passo di Levesque e Brachman sopra citato. Un sistema logico-formale è caratterizzato da un *linguaggio*, da un insieme di *assiomi* e da un insieme di *regole di inferenza*. Il linguaggio corrisponde alle strutture del sistema di rappresentazione della conoscenza. Perché si tratti di un sistema logico, alle espressioni che appartengono al linguaggio deve essere possibile associare una semantica di tipo formale. Le regole di inferenza sono l'analogo dei meccanismi di inferenza del sistema di rappresentazione della conoscenza: esse devono consentire di elaborare le formule del sistema formale esclusivamente sulla base alle loro caratteristiche sintattiche, ignorandone cioè il significato. Tuttavia le trasformazioni ottenute devono essere in accordo con la semantica, con la teoria della verità, del sistema formale.

Senza dubbio, perché ciò abbia senso, bisogna adottare un concetto esteso di logica, non limitato ai sistemi logici e alla notazione tradizionale. Moore (1982) prende questa direzione. Egli analizza alcuni aspetti della logica dei predicati, come ad esempio il fatto di rappresentare il mondo nei termini di individui, proprietà e relazioni, e la possibilità di rappresentare conoscenza incompleta per mezzo di disgiunzione, negazione e quantificazione esistenziale. Egli quindi afferma: "Mi sembra che, nella misura in cui qualunque formalismo di rappresentazione ha le caratteristiche logiche discusse sopra, esso è una logica, e che, nella misura in cui una procedura di ragionamento prende in considerazione tali caratteristiche, essa ragiona deduttivamente. E' concepibile che possa esistere un modo di trattare questi aspetti che sia radicalmente differente dalle logiche correnti, ma esse sarebbero ancora *qualche* tipo di logica" (pp. 430-31).

Dopo i violenti dibattiti degli anni '70, a partire dalla prima metà degli anni '80 la linea logicista è quella che ha raccolto i maggiori consensi fra i ricercatori di IA simbolica¹⁶. Questo non implica né che si sia raggiunto un totale consenso sulla possibilità di ridurre tutta l'IA di tipo simbolico al paradigma logicista, né che in questo ambito possano trovare una soluzione tutti i problemi posti dall'IA¹⁷. Tuttavia, il filone logicista è stato innegabilmente il più sistematico ed il più ricco di risultati.

Tornando al tema della semantica cognitiva e della rappresentazione del significato, gli assunti della linea logicista in IA comportano che le rappresentazioni costruite nel corso della comprensione di enunciati linguistici vengano identificate con espressioni di un linguaggio logico (nel senso lato visto prima), o almeno possano essere studiate con gli strumenti della logica matematica - il che significa soprattutto che deve essere possibile associare loro una semantica di tipo modellistico. Quindi, come avevamo anticipato, questo settore della scienza cognitiva nega esplicitamente il punto 5. della definizione di Marconi sopra riportata. Il logicismo in AI sembra dunque volere conciliare in qualche modo l'idea di significato come costruzione "mentale" e la tesi del significato definito nei termini di condizioni di verità (in questo caso, ciò che diventa rilevante sono le condizioni di verità delle costruzioni mentali). Bisognerebbe ovviamente precisare cosa si debba intendere per "costruzione mentale" in questo contesto. E' infatti evidente per l'IA di impostazione logica non è rilevante l'adeguatezza psicologica in senso empirico delle strutture di rappresentazione ipotizzate. Non si può quindi parlare di rappresentazioni mentali nel senso che interesserebbe uno psicologo (non si tratta in senso stretto di una semantica psicologica). Tuttavia si ha a che fare comunque con costruzioni elaborate da soggetti razionali (eventualmente artificiali), o da modelli di soggetti razionali, comunque individuali e finiti. Si tratta delle rappresentazioni (o di modelli delle rappresentazioni) che stanno "dentro la testa" (o "dentro la memoria") di un singolo agente - e quindi comunque non conciliabili con le assunzioni antipsicologiche di

¹⁶ Per convincersene, basta sfogliare gli atti dei più importanti convegni internazionali.

¹⁷ Si veda ad esempio (McDermott 1987) per una critica al logicismo non sospettabile di faziosità, in quanto dovuta ad uno dei maggiori rappresentanti di questa stessa linea.

Frege e della tradizione model teoretica classica in filosofia del linguaggio. Torneremo su questi problemi nelle conclusioni.

Per il momento, si tratta innanzi tutto di capire quale contributo l'IA logicista possa fornire alla soluzione dei problemi posti dalla semantica modellistica. Nelle due prossime sezioni esamineremo quindi rispettivamente i formalismi e i metodi proposti per la rappresentazione del significato lessicale, e le conseguenze per il problema di una rappresentazione del significato dei simboli primitivi extra logici (capp. 4-7), e le soluzioni proposte al problema degli atteggiamenti proposizionali, in particolare dell'onniscienza logica (capp. 8-11). Prima di procedere, sono doverose due precisazioni. La prima concerne il fatto che qui ci occuperemo principalmente di *rappresentazione* del significato, e non della generazione di tale rappresentazione a partire dagli *input* in linguaggio naturale. Quindi, la nostra trattazione avrà a che fare con aspetti che riguardano il settore della *rappresentazione della conoscenza* piuttosto che il settore dell'elaborazione del linguaggio naturale. Questo non esclude ovviamente che esista un grande numero di temi di elaborazione del linguaggio naturale che risulterebbero estremamente rilevanti per il tema qui trattato. Tuttavia, non potendo ovviamente esaurire l'intero settore, riteniamo che i problemi che ci interessano emergano in maniera più esplicita ed evidente in rappresentazione della conoscenza¹⁸.

La seconda precisazione concerne il fatto che le due sezioni sulla rappresentazione del significato lessicale e sul problema dell'onniscienza logica sono per molti versi asimmetriche. Tale asimmetria deriva dal fatto che il primo problema non è tematizzato in maniera specifica nella IA logicista (anche perché, come abbiamo visto, gli strumenti logici e model teoretici in quanto tali non consentono neppure in linea di principio una soluzione definitiva al problema della rappresentazione del significato dei simboli primitivi extra logici di un linguaggio). In questa sezione quindi si è tentato di caratterizzare l'evoluzione dalle proposte di parte anti logica per il problema della rappresentazione concettuale e lessicale (soprattutto reti semantiche e *frame*) fino all'egemonia del logicismo. Viceversa, il ragionamento sul credere è un settore ben delimitato della IA di impostazione logica. Nella seconda parte si è cercato quindi di fornire un quadro delle ricerche che concernono il problema dell'onniscienza logica in ambito logicista, tralasciando i lavori sullo stesso argomento esterni a questa tradizione¹⁹. Di conseguenza, la sezione sull'onniscienza logica risulta molto più omogenea di quella sulla rappresentazione lessicale, che spazia fra aree di ricerca molto diversificate. Questo ha fatto sì che nella sezione sull'onniscienza logica si sia potuto aspirare, se non a una completa esaustività, almeno a un resoconto abbastanza dettagliato delle varie proposte; viceversa, nella prima sezione è stato possibile fornire soltanto una panoramica di ampio respiro, effettuando alcune scelte motivate in parte da decisioni personali, ma comunque tali da mettere in luce le questioni di fondo del problema.

¹⁸Un altro tema che non affronteremo in questo lavoro, anche se presenterebbe motivi di interesse per l'indagine dei rapporti fra IA logicista e teorie del significato, è costituito dai lavori concernenti la formalizzazione di aspetti *pragmatici* del linguaggio, come ad esempio i vari tentativi di formalizzare gli atti linguistici o il dialogo. Inoltre, sempre per ragioni di spazio, non prenderemo in considerazione quelle teorie "eterodosse" del significato nate dal filone della semantica modellistica, come ad esempio la *situation semantics* di Barwise e Perry (1983) o la teoria di Kamp (sul problema degli atteggiamenti proposizionali nella teoria di Kamp si veda Castelnovo 1986), che pure si sono spesso sviluppate secondo una prospettiva di tipo computazionale.

¹⁹Come ad esempio i lavori di Maida e Shapiro (Maida e Shapiro 1982; Maida 1985).