

# Gli agenti dall'informatica alle scienze cognitive e alle applicazioni

### 6.1. I confini della metodologia

Uno dei problemi della simulazione ad agenti è la definizione dei confini della metodologia informatica sottostante. Dal punto di vista dell'ingegneria del software, un sistema di programmi indipendenti (eventualmente interagenti), è detto *multi-agente*. Tuttavia, non vi è chiarezza e univocità su ciò che il termine «agente» definisce e sottende; ci si propone qui di analizzare le varie accezioni di questo concetto e tracciare, per quanto possibile, una semplice tassonomia degli agenti software propriamente detti. In particolare, l'obiettivo è quello di identificare quali tipologie di agenti possano essere utilizzate con successo per modellare e simulare sistemi sociali complessi. Nella trattazione che seguirà, sarà introdotto e spesso utilizzato il concetto di *Multi Agent System* (MAS); in generale sarebbe possibile definire MAS tutte le situazioni in cui più entità autonome operano all'interno dello stesso ambiente.

Esiste tuttavia una definizione più forte: un MAS è una rete di agenti software che interagiscono tra loro per risolvere problemi e raggiungere obiettivi che vanno oltre le capacità e le conoscenze individuali di ognuno di essi. In questo particolare contesto si preferisce tale definizione a quella di *Agent Based System* (ABS) semplicemente per enfatizzare la presenza di più agenti che interagiscono, poiché, nel corso del capitolo, si farà riferimento anche a sistemi che possono essere «monoagente».

Il termine agente deriva dal latino *agens*, vocabolo che identifica qualcuno (o qualcosa) che agisce; in un'accezione più ampia è possibile utilizzare il termine per identificare un mezzo attraverso cui qualche azione è compiuta o causata. Gli esseri umani agiscono e così la maggior parte degli animali; sulla base della definizione più ampia, anche degli oggetti inanimati possono essere considerati come agenti. Il termine si incontra infatti in discipline molto diverse tra loro, quali economia, fisica, scienze naturali, sociologia e molte altre. In ambito informatico, tuttavia, il vocabolo è utilizzato per definire entità eterogenee tra loro e spesso se ne abusa, andando oltre al concetto identificato sopra. Ecco dunque che talvolta dei semplici oggetti, cioè parti di codice informatico scritto in

linguaggio *object oriented* sono definiti agenti, nonostante non abbiano le proprietà che si vorrebbero proprie dei veri agenti software.

Per evitare ciò, in Franklin e Graesser [1996] troviamo alcuni precetti che aiutano a definire i confini dei concetti di «agente» ed «agenzia» nell'ambito informatico; i due autori notano che in Wooldridge e Jennings [1995] un agente è definito come un sistema hardware o – più spesso – software che presenta le seguenti proprietà: *autonomia*, cioè capacità degli agenti di operare senza l'intervento diretto degli umani o di altri dispositivi e per questo dotato di un qualche meccanismo di controllo delle proprie azioni; *abilità sociale*, cioè la capacità degli agenti di interagire con altre entità (siano esse altri agenti software, esseri umani o altro) attraverso un qualche tipo di linguaggio di comunicazione; *reattività*, cioè la possibilità di rispondere in modo coerente e sincronizzato agli stimoli provenienti dall'ambiente in cui sono immersi – che può essere costituito dal mondo fisico oppure da una simulazione; *pro-attività*: la capacità di agire sulla base di uno scopo, un obiettivo che può essere definito dal programmatore o creato internamente ed autonomamente dall'agente, partendo da una serie di elementi logici presenti in esso.

Norman [1994] osserva che i predecessori più significativi degli agenti intelligenti di oggi sono i servomeccanismi ed altri sistemi di controllo, inclusi quelli utilizzati nelle catene di montaggio e quelli tipici degli aerei, come i sistemi di decollo e atterraggio automatici e di pilotaggio e volo automatico. Tuttavia, le entità attualmente definite «agenti» differiscono per molti aspetti dalla concezione primitiva; in particolare, almeno per il momento, l'accento pare essersi spostato dall'hardware al software. Il concetto di agente ha origine negli anni '50 con John McCarthy, mentre il termine fu coniato da Oliver G. Selfridge qualche anno più tardi, quando entrambi lavoravano al Massachusetts Institute of Technology. Essi avevano come progetto un sistema che, dato un obiettivo da compiere, fosse in grado di progredire, chiedendo eventualmente aiuto agli umani in caso di stallo o di mancanza di informazioni. Un agente era immaginato come un robot software che vive ed opera all'interno di un mondo virtuale.

Nwana [1996] suddivide gli studi sugli agenti software in due filoni principali: il primo, con inizio nel 1977, si fonda sugli studi nel campo dell'intelligenza artificiale distribuita e si è concentrato soprattutto sugli agenti deliberativi con modelli simbolici interni. Il secondo, le cui origini risalgono ai primi anni '90, ha come oggetto una vasta gamma di tipologie di agenti, e l'enfasi qui è posta sul fare, anziché sul decidere.

Con la nascita di tipi di agenti software eterogenei e molto diversi tra loro, il termine non è sufficiente per identificare il campo di applicazione; certi codici informatici sono definiti «agenti» solamente in quanto possono essere programmati per compiere operazioni su una macchina remota; alcuni poiché compiono operazioni di basso livello, ma sono programmati con linguaggi o script di alto livello; altri perché astraggono ed incapsulano i dettagli delle differenze tra sorgenti di informazione; altri ancora perché implementano una «funzione cognitiva» aggregata [Minsky, 1986]. Sono anche definiti agenti quei programmi che manifestano caratteristiche di intelligenza distribuita, come si legge in Moulin e Chaib-Draa [1996]; quelli che servono come mediatori tra persone e programmi [Coutaz, 1990; Wiederhold, 1989], o rivestono il ruolo di «assistenti artificiali» [Boy, 1991; Maes, 1997] e ancora quelli che possono migrare da un computer all'altro autonomamente [White, 1996]. In Shoham [1993] l'agente è definito come un'entità il cui stato consiste di componenti mentali quali credenze, capacità, scopi e possibilità di

operare scelte. Etzioni [1993] presenta invece gli agenti come robot software (*softbots*) dotati di sensori che possono variare il loro stato in base all'ambiente esterno.

La definizione probabilmente più generale e capace di catturare l'essenza del concetto si trova in Jennings [1996], dove si legge che il termine è solitamente utilizzato per descrivere programmi o parti di essi in grado di controllare le proprie azioni, basandosi sulle percezioni provenienti dall'ambiente in cui operano.

Ancora Shoham [1997] fornisce una caratterizzazione più specifica, descrivendo un agente come un'entità software che funziona continuamente ed autonomamente in un particolare contesto, spesso popolato da altri agenti e processi. La richiesta di continuità ed autonomia deriva dal desiderio che un agente sia in grado di svolgere attività in modo flessibile ed intelligente e sia reattivo nei confronti dei cambiamenti dell'ambiente senza richiedere una guida ed un intervento costante da parte di umani. Idealmente, un agente che funzioni in modo continuo all'interno di un ambiente per un lungo periodo di tempo potrebbe essere in grado di apprendere dalla propria esperienza. Inoltre ci si aspetta che un agente che convive con altri agenti e processi possa comunicare e cooperare con essi e, in caso di necessità, anche muoversi da un luogo all'altro all'interno del proprio ambiente. Questi sono solo alcuni esempi di entità software che possono essere classificate come agenti; nessuna di queste definizioni è scorretta, ma ci si riferisce a tipi di agenti molto diversi tra loro.

## 6.2. Caratteristiche fondamentali degli agenti

Anche se non esiste una definizione univocamente accettata, possiamo utilizzarne alcune per definire le caratteristiche fondamentali degli agenti intelligenti ed in particolare per identificare le categorie che possono essere utilizzate con successo per lo studio di sistemi sociali complessi.

Dalla lettura comparata di Brustoloni [1991] e Franklin [1995] possiamo dedurre che gli agenti cosiddetti autonomi devono essere in grado di compiere azioni dotate di uno scopo; anche se Brustoloni si riferisce in particolare ad agenti hardware, che agiscono ed operano nel mondo reale, questo concetto può essere applicato con successo anche a quelli software, presenti all'interno di una simulazione, come quelli citati da Langton [1989], il quale specifica che «gli agenti dotati di vita artificiale vivono in ambienti virtuali su uno schermo di computer e all'interno della sua memoria».

Gilbert *et al.* [1995] descrivono gli agenti intelligenti nei termini di uno spazio definito da tre dimensioni: *agency*, intelligenza e mobilità. L'*agency* è il grado di autonomia ed autorità rivestita dall'agente e può essere misurata almeno qualitativamente dalla natura dell'interazione tra l'agente e le altre entità nel sistema. Il livello di *agency* risulta più elevato se un agente rappresenta un utente in qualche modo. L'intelligenza è il grado di ragionamento e di comportamento appreso: l'abilità che un agente dimostra nell'interpretare correttamente e portare a termine i compiti che gli sono affidati dall'utente. La mobilità è la caratteristica per la quale un agente può muoversi su una rete; se collegata all'intelligenza prima descritta, la mobilità permette ad un agente di decidere se e quando muoversi sulla rete, in accordo con le esigenze dettate dai compiti da svolgere.

Una definizione generale, ma efficace, del concetto di agente si trova in Franklin e Graesser [1996]: «un agente autonomo è un sistema inserito in – e parte di – un ambiente, in grado di percepire ciò che lo circonda ed agire di conseguenza, per portare

a termine determinati obiettivi e perseguire ciò che esso programma per il futuro». Da questa definizione si evince che un agente software può avere una proiezione dello stato del mondo attuale, in base alla quale deciderà quali azioni compiere, degli obiettivi, derivanti dal modo in cui è costruito, e delle aspettative sul futuro; in altri termini deve essere in grado di elaborare una catena di cause ed effetti sulla base delle azioni che deciderà di compiere.

Si può distinguere tra una definizione debole di *agency* ed una forte; la prima comprende le proprietà prima elencate, alle quali si aggiungono il funzionamento asincrono rispetto agli altri agenti e la continuità temporale, fondamentale per una rappresentazione realistica di un ambiente virtuale in cui vi sia interazione tra più entità. In accordo con questa definizione debole di *agency* un agente, dal punto di vista informatico, può essere considerato come un elemento software auto-contenuto (che cioè contiene tutto ciò che è necessario per il proprio funzionamento) e con esecuzione parallela, che possiede degli stati interni e comunica con altri agenti attraverso lo scambio di messaggi, sviluppato con l'utilizzo di un linguaggio *object oriented*.

Esiste anche una nozione più forte e specifica di *agency* dal punto di vista informatico; un agente è dunque definito come un sistema che, in aggiunta alle proprietà precedentemente elencate, è progettato o realizzato utilizzando concetti che normalmente sono applicati alla descrizione di esseri umani, quali l'intelligenza. In questa accezione si può dunque parlare di cooperazione tra agenti e di apprendimento basato sui precedenti; in particolare, il primo concetto è applicato con notevoli risultati ai sistemi multi agente (MAS).

### 6.3. Struttura di un sistema multi agente

Un sistema multi agente ha numerosi vantaggi nei confronti di un modello ad agente singolo o centralizzato.

- Distribuisce le risorse e capacità computazionali su di una rete di agenti interconnessi. Un MAS è tipicamente decentralizzato e pertanto non è caratterizzato dall'esistenza di un punto cruciale unico, riscontrabile in sistemi centralizzati.
- Permette l'interconnessione e l'interoperabilità di più sistemi già esistenti, che possono essere incorporati e visti a loro volta come agenti, all'interno del nuovo MAS.
- Modella i problemi in termini di componenti elementari che interagiscono tra loro; questo è un modo molto naturale di rappresentare le situazioni, soprattutto nei sistemi sociali.
- Raccoglie, filtra e coordina informazioni provenienti da sorgenti distribuite spazialmente.
- Fornisce soluzioni in situazioni dove l'esperienza è distribuita nello spazio e nel tempo.

- Incrementa le prestazioni globali del sistema, specificatamente per quanto riguarda l'efficienza computazionale, l'affidabilità, l'estensibilità, la robustezza, la rapidità di risposta e la flessibilità.

Gli agenti facenti parte di un MAS possono interagire positivamente o negativamente; nel primo caso si parla di cooperazione, nel secondo di competizione. Una caratteristica fondamentale dei MAS è che a livello aggregato essi possono mostrare delle proprietà che non appartengono ad alcuno degli agenti che lo formano; questa sorta di sinergia, che permette un'emergenza a livello macro di nuove proprietà, permette all'occorrenza di studiare il MAS come un agente unico. Le proprietà tipiche di un MAS sono:

- il modello del sistema che si intende studiare;
- la granularità (cioè il dettaglio);
- il numero degli agenti;
- la capacità dei singoli agenti e del sistema intero di adattarsi (che può essere fissa, programmabile, con apprendimento e così via);
- la distribuzione e il grado del controllo (dipendente o indipendente dall'utente umano);
- la quantità delle risorse facenti capo ai singoli agenti (limitate o illimitate);
- lo schema di interazione tra gli agenti;
- il grado di cooperazione tra agenti.

Al fine di massimizzare l'efficienza del sistema ogni agente dev'essere in grado di compiere una sorta di ragionamento riguardante le azioni che possono essere portate a termine dalle altre entità, oltre che gli effetti delle azioni proprie. Un ambiente dinamico ed imprevedibile porta alla necessità, da parte degli agenti, di adottare strategie flessibili. Tuttavia, con l'aumento della flessibilità diviene più difficile prevedere come gli agenti si comporteranno, date determinate situazioni; per questo motivo esistono meccanismi di coordinamento che servono per aiutare gli agenti a livello aggregato, quando essi si trovano a compiere azioni complesse che richiedono un lavoro di gruppo per essere portate a termine. Questi meccanismi devono assicurare che i piani degli agenti individuali non vadano in conflitto tra di essi, mentre perseguono i macro obiettivi richiesti dal sistema.

Tradizionalmente, come si legge in Wooldridge e Jennings [1995] è possibile raggruppare gli agenti software in tre categorie fondamentali:

- reattivi;
- deliberativi;
- ibridi.

Quando si desidera costruire un sistema basato su agenti software, è necessario determinare quanto sofisticate dovranno essere le capacità di ragionamento delle entità coinvolte; gli agenti reattivi sono in grado semplicemente di richiamare comportamenti predefiniti, metaforicamente simili ai riflessi ed agli istinti umani, senza tener conto di stati interni che giustifichino diverse modalità di funzionamento. Al contrario gli agenti deliberativi si comportano come se fossero in grado di pensare, ponendo in essere piani composti dall'unione di molti comportamenti elementari, tenendo conto del proprio stato interno e soprattutto cercando di prevedere gli effetti delle proprie azioni, prima di compierle. Anche se la linea di confine tra i due paradigmi, nella pratica, può spesso essere poco marcata, possiamo dire che un agente che non ha uno stato interno mutevole è sicuramente di tipo reattivo, mentre quello che basa le proprie azioni sulle previsioni delle conseguenze e del comportamento degli altri agenti è sicuramente di tipo deliberativo.

#### 6.4. Azione e reazione

Gli agenti reattivi rappresentano una categoria speciale di agenti, privi di modelli interni simbolici dell'ambiente in cui operano: agiscono in risposta agli stimoli che provengono dall'esterno. Le ricerche sugli agenti reattivi iniziarono con Brooks [1986] ed Agre e Chapman [1987], ma da allora sono state sviluppate molte teorie, architetture e linguaggi per la loro rappresentazione. Tali agenti hanno solitamente un'architettura molto semplice ed interagiscono con altri spesso in modo elementare; tuttavia, da queste interazioni semplici possono emergere comportamenti complessi (*patterns*), quando si considera l'aggregato anziché i singoli agenti.

Maes [1991] indica le tre idee chiave che distinguono gli agenti reattivi; in primo luogo, la *funzionalità emergente*, cioè il comportamento aggregato complesso che ha origine da un insieme di semplici azioni elementari, anche quando non vi è alcuna specifica o piano a priori. La seconda importante caratteristica è definita *scomposizione dei compiti*: un agente reattivo è visto come un insieme di moduli che operano autonomamente e sono responsabili di compiti specifici (percezioni, controllo del movimento, calcolo e così via). Non esiste un modello globale per alcuno degli agenti e, pertanto, il comportamento globale deve emergere. Infine gli agenti reattivi tendono ad *operare su rappresentazioni* che sono simili a dati grezzi, il che li pone in contrapposizione con le rappresentazioni simboliche di alto livello presenti, per esempio, negli agenti di tipo deliberativo. In Mataric [1995] si legge che gli agenti reattivi non possiedono un modello interno con cui predire i futuri stati del mondo; scelgono le azioni da compiere utilizzando lo stato attuale del mondo come indice in una tavola di possibili azioni, in cui lo scopo della funzione indice è quello di collegare situazioni note verso le azioni appropriate.

Questo tipo di agenti è sufficiente per ambienti limitati, dove ogni possibile situazione può essere prevista e collegata ad un insieme di azioni conseguenti. Il maggior difetto degli agenti di tipo reattivo è la carenza di adattabilità; infatti essi non possono generare autonomamente piani nel caso in cui un determinato stato del mondo non sia considerato dal loro programma. Nei domini che non possono essere completamente interconnessi, dunque, l'uso di agenti puramente reattivi può essere troppo limitante.

In Balch e Arkin [1995] si utilizzano agenti omogenei, reattivi e privi di forme di comunicazione tra loro per studiare formazioni militari; l'obiettivo dato agli agenti è quello

di muoversi insieme, mantenendo una predeterminata formazione (a diamante, colonna, ...), senza appunto comunicare. Possono, inaspettatamente, incontrare ostacoli sul loro cammino tali da impedire ad uno o più agenti di muoversi in linea retta. Dopo aver oltrepassato l'ostacolo, tutti gli agenti devono correggere la propria posizione per ritornare alla formazione originaria. Per ottenere questo risultato gli agenti, in modo reattivo, convertono i dati provenienti dai loro sensori (posizione propria, posizione degli altri agenti, conformazione del terreno, ...) in vettori di moto che consentano loro di evitare ostacoli e altri agenti, muovendosi verso una posizione obiettivo, che permetta, a livello aggregato, di mantenere la formazione.

### 6.5. Ragionamento negli agenti

A differenza degli agenti reattivi, quelli deliberativi hanno, come componente chiave, un sistema di ragionamento [Ginsberg, 1989] che costituisce l'intelligenza dell'agente. Gli agenti di tipo deliberativo sono in grado di generare piani per portare a termine i propri obiettivi; possono avere un modello del mondo al loro interno e questo accresce la loro capacità di generare piani in maniera razionale, tenendo conto delle relazioni di causa ed effetto.

Il maggior pregio di questo paradigma consiste nell'essere in grado di affrontare situazioni impreviste in ambienti di tipo dinamico; il problema è invece costituito dal tempo necessario per elaborare piani coerenti, spesso troppo elevato per sistemi operanti in tempo reale, in cui è necessario prendere decisioni rapidamente. Wooldridge [1995] definisce agente deliberativo quello che possiede un modello del mondo rappresentato esplicitamente in cui le decisioni – per esempio quali azioni compiere – sono assunte attraverso un ragionamento simbolico.

Inizialmente il lavoro su questo tipo di agenti si concentrò su problemi a livello macro, come la comunicazione e l'interazione tra agenti, la scomposizione e distribuzione dei compiti, il coordinamento e la cooperazione, la risoluzione di conflitti attraverso la negoziazione e così via. Come esempio di agenti deliberativi si citano Levy e Rosenschein [1992]; gli agenti da loro progettati partono dal presupposto che ognuno di essi agirà per portare a termine gli obiettivi assegnati. Utilizzano tecniche derivate dalla teoria dei giochi per trovare punti di equilibrio e decidere come agire. Questi agenti sono chiaramente deliberativi, in quanto ricercano le azioni da compiere sulla base di funzioni di massimizzazione, piuttosto che richiamarle semplicemente da una banca dati.

### 6.6. Soluzioni ibride

Gli agenti di tipo ibrido, se progettati correttamente, utilizzano entrambe le tecniche viste precedentemente per ovviare agli inconvenienti citati. In Bensaïd e Mathieu [1997] si legge che gli agenti di tipo ibrido mirano ad avere un tempo di risposta rapido, tipico degli agenti reattivi, quando si affrontano situazioni note, ma anche la capacità di generare nuovi piani quando si affrontano casi non previsti.

Esistono diversi sistemi e tecniche che uniscono comportamenti reattivi e deliberativi; un esempio è quello che si trova in Rao e Georgeff [1995], dove gli agenti possono scegliere, sulla base delle situazioni che incontrano, se essere reattivi – dunque istintivi,

con minor tempo di reazione, ma anche minor precisione nell'azione – oppure se costruire veri e propri piani di azione, utilizzando il paradigma deliberativo, in casi in cui ciò sia considerato necessario. Tali architetture sono considerate ibride, anche se alla base prevale uno dei due paradigmi; nell'esempio analizzato, in particolare, gli agenti sono fundamentalmente di tipo deliberativo, in quanto alla base della scelta su come operare (in modo reattivo o meno) c'è comunque un ragionamento sulle situazioni che si incontrano.

I sistemi ibridi possono essere molto efficienti in quanto, in casi noti o prevedibili, l'agente può agire con tempi di reazione molto bassi, tipici dell'architettura reattiva, mentre per situazioni mai incontrate o complesse sarà in grado di utilizzare algoritmi di intelligenza artificiale per simulare un ragionamento e portare a termine azioni appropriate, anche quando queste non fossero state esplicitamente previste dal progettista del sistema.

Un altro tipo di sistema ibrido è quello descritto in Sahota [1994], con la tecnica di deliberazione reattiva. In questo caso un agente ragiona, in modo deliberativo, su quale comportamento tenere, sotto il vincolo che le proprie azioni devono essere scelte con una frequenza predefinita. Questo tipo di agente è adatto alle situazioni in cui la maggior criticità sta nei tempi di reazione e quando è preferibile assumere una decisione non ottimale, piuttosto che non scegliere in tempo. Anche se tale paradigma non è stato originariamente pensato per i sistemi multi agente, può essere adottato anche in quel campo; per esempio, è possibile che alcuni degli agenti di un MAS, anche se dotati di sistemi di intelligenza artificiale, debbano decidere – o agire – pur non avendo portato a termine il loro processo decisionale, magari per essere sincronizzati con altri agenti dello stesso sistema. In questo caso può essere utile dotare tali agenti di un sistema di ottimizzazione parziale, con una soglia temporale prefissata, rendendoli in grado di effettuare l'azione migliore possibile in un determinato tempo.

Se gli agenti presentano connessioni tra loro, si utilizzano normalmente due diverse tipologie di legami.

- Connessioni permanenti, predeterminate progettando il sistema. Queste connessioni sono rappresentabili come linee continue in un grafo e mostrano prevalentemente relazioni di comunicazione tra gli agenti. Sono immutabili nel tempo e pertanto non se ne possono estinguere o creare di nuove.
- Connessioni *ad hoc*, cioè non predeterminate, create dinamicamente durante l'esecuzione del sistema. Solitamente queste connessioni sono graficamente rappresentate attraverso linee tratteggiate e possono variare a seconda delle esigenze contingenti del sistema.

Solitamente le connessioni del primo tipo sono utilizzate quando gli agenti sono di tipo reattivo, mentre per quelli di tipo deliberativo è preferita la seconda tipologia.

### 6.7. Agenti BDI, scienze cognitive e logica modale

Passando ora al campo non solo nell'ambito delle discipline informatiche, ma a quello delle scienze cognitive, introduciamo una importante classe di agenti di tipo deliberativo: quella degli agenti BDI (*Beliefs, Desires, Intentions*); si tratta di un paradigma di agenti sviluppato principalmente dal punto di vista teorico, in quanto l'applicazione pratica risulta quasi irrealizzabile, a meno di non rinunciare a caratteristiche peculiari. Bratman [1987], in un lavoro che si colloca tra filosofia ed intelligenza artificiale, per primo individua come componenti chiave dell'agire le credenze, i desideri e gli obiettivi; da allora, i formalismi che danno un'importanza primaria alle intenzioni, sono definiti architetture BDI.

Mentre le teorie precedenti trattano le intenzioni come se fossero riconducibili alle credenze ed ai desideri, Bratman sostiene che le intenzioni esercitano un ruolo distinto nel ragionamento pratico. Le intenzioni sono così considerate come piani parziali per l'azione che l'agente deve eseguire per raggiungere i propri scopi. Cohen e Levesque [1990] definiscono le azioni in termine di sequenze temporali delle credenze e obiettivi dell'agente.

Ricordiamo anche che molte tra le caratteristiche tipiche degli agenti provengono dagli studi effettuati nel campo delle scienze cognitive; con Legrenzi [2002] ricordiamo che scienza cognitiva spiega tra l'altro i modi in cui menti naturali o artificiali filtrano e colgono informazioni percettive, le rielaborano e riescono a intraprendere delle decisioni in base alle circostanze, tanto da reagire al mondo esterno anche elaborando degli artefatti. Le scienze cognitive sono la sintesi di risultati che provengono dall'antropologia, dalla logica, dalla filosofia, dalla psicologia, dalla linguistica e dalle neuroscienze, nello studio delle facoltà mentali. Nascono negli anni '70 del secolo scorso, grazie al confluire dei risultati di più discipline e hanno un carattere fortemente interdisciplinare.

In particolare i primi modelli teorici di agenti, tenendo conto degli studi in questo campo sul funzionamento della mente, tendevano a mimare il comportamento cognitivo degli esseri umani, essendo orientati verso obiettivi precisi per ottenere i quali potevano tentare di compiere ragionamenti deliberativi, apprendere, comunicare tra loro e così via.

Tuttavia le implementazioni pratiche, dovendo affrontare problemi gravati dall'enorme complessità computazionale, hanno dovuto sacrificare parte di quelle caratteristiche, concentrandosi maggiormente su aspetti specializzati, anche se sempre derivati dall'intelligenza, con agenti apparentemente più semplici, ma più reattivi, posti anche alla base dei modelli di tipo sociale.

### 6.8. I mondi possibili e la logica modale

Nel lavoro di Rao e Georgeff [1991] le architetture BDI sono per la prima volta definite attraverso il formalismo dei mondi possibili e della logica modale. La logica modale è una forma di logica che tratta proposizioni qualificate da caratteristiche di modalità quali: possibilmente, necessariamente, contingentemente, può, potrebbe e così via. A differenza dalle forme tradizionali di logica del prim'ordine, che può solamente operare su asserzioni, siano esse vere o false (quali, per esempio «Socrate è mortale», o «tutti i cani sono dei rettili»), la logica modale tratta anche relazioni logiche tra affermazioni

non certe, quale «è possibile che giovedì piova» e frasi apodittiche come «se si sommano 2 e 3 la somma deve necessariamente essere 5». Si ritiene *necessaria* una proposizione che deve per forza essere vera; *impossibile* una proposizione che deve per forza essere falsa; *contingente* una proposizione che non è né necessaria né impossibile (alcune proposizioni contingenti saranno talvolta vere e talvolta false); *possibile* una proposizione che non è impossibile (in questa classe sono perciò incluse tutte le proposizioni, eccetto quelle impossibili). Questi quattro aggettivi si denotano con il termine, derivato dalla tradizione logica medievale, di connettivi modali o modalità, in quanto si ha a che fare con i «modi» in cui una proposizione può essere vera; premettendo uno qualsiasi di questi connettivi ad una proposizione si ottiene una nuova proposizione.

Nell'interpretazione più comunemente accettata si considerano «tutti i mondi possibili» per definire proposizioni di logica modale, cioè tutte le situazioni immaginabili; se una proposizione è vera in tutti i mondi possibili, allora è necessariamente vera. Se una proposizione è vera nel mondo in cui ci si trova, ma non vera in tutti i mondi possibili, allora è contingente; una proposizione vera in qualche mondo possibile (non necessariamente in quello in cui ci si trova) è detta possibile.

Tre sono gli elementi fondamentali per questa teoria: in primo luogo le intenzioni rivestono la stessa importanza dei desideri e degli obiettivi. Inoltre si distingue tra la scelta che un agente può operare tra le azioni da compiere e le conseguenze possibili di una stessa azione. Nel primo caso, infatti, è l'agente a scegliere, mentre nel secondo è l'ambiente a determinare l'effetto. Infine sono specificate relazioni – tra credenze, obiettivi ed intenzioni – che permettono di evitare problemi classici legati ai formalismi dei mondi possibili, come seguire qualche effetto collaterale anziché l'obiettivo vero e proprio.

Per modellare le azioni degli agenti BDI si utilizza una struttura temporale ad albero, orientato verso il futuro; un particolare istante in un determinato mondo è detto situazione. Gli eventi trasformano un punto in un altro ed i rami dell'albero sono una metafora per identificare le scelte possibili per gli agenti. Per esempio, se da un certo punto dell'albero si dipartono due rami, uno definito E1 e l'altro E2, allora l'agente ha possibilità di scegliere quale dei due percorrere, cioè, metaforicamente, quale azione compiere.

Ovviamente un agente potrebbe decidere di compiere una certa azione, ma non riuscire a svolgerla: in questo caso il ramo verrebbe contrassegnato come azione fallita. Il formalismo usato è derivato dalla *Tree Logic* di Emerson *et al.* [1998]; si introducono due operatori modali, «opzionale» ed «inevitabile», per contrassegnare le formule logiche che identificano un dato sentiero. Una formula che definisce un sentiero è detta opzionale se, in un particolare punto dell'albero temporale, la formula è vera almeno per un sentiero che si diparte da quel punto; è invece detta inevitabile se è vera per tutti i sentieri che si dipartono dal punto. Si utilizzano quindi gli operatori temporali considerati standard, cioè: «prossimo», «prima o poi», «sempre», «fino a». Per esempio, grazie alla combinazione degli operatori appena citati è possibile creare proposizioni logiche quali: «è opzionale che prima o poi Mario visiterà Londra» e «è inevitabile che uno più uno dia sempre due come risultato».

Le credenze sono così modellate: ad ogni situazione è associato un insieme di mondi accessibili, che l'agente ritiene essere possibili. Ognuno di questi mondi è un albero

temporale e all'interno di ognuno di questi mondi le ramificazioni rappresentano le scelte ancora disponibili che l'agente può compiere. In modo simile, ad ogni situazione si associa anche un insieme di mondi che rappresentano gli obiettivi dell'agente; mentre i desideri possono non essere realizzabili, gli obiettivi sono desideri selezionati dall'agente, sicuramente realizzabili considerando la situazione contingente. In questo modo si impedisce agli agenti di perseguire obiettivi ritenuti sin dall'inizio irrealizzabili; questa proprietà è definita realismo [Cohen e Levesque, 1987] ed è rinforzata da Rao e Georgeff [1991] attraverso la richiesta che, per ogni mondo-credenza  $w$  accessibile in un certo momento  $t$ , ci debba essere un mondo-obiettivo che sia un sottomondo di  $w$ . Anche le intenzioni sono rappresentate come un insieme di mondi accessibili: sono infatti quelli che l'agente ha deciso di provare a realizzare. Tali mondi devono essere compatibili con i mondi-obiettivo, cioè un agente può avere l'intenzione di percorrere una certa azione solo se è uno dei propri obiettivi. Dunque, in corrispondenza di un dato mondo-obiettivo  $w$  all'istante  $t$ , ci dev'essere un mondo-intenzione che sia un sottomondo di  $w$ . Intuitivamente, l'agente sceglie un corso per le proprie azioni tra quelle ritenute realizzabili in  $w$  e si propone di tentarne l'esecuzione. In questo ambito i diversi mondi che rappresentano credenze, obiettivi ed intenzioni costituiscono scenari possibili per gli agenti; l'agente ritiene che il mondo attuale, cioè quello in cui è immerso, sia uno dei mondi possibili.

Boella *et al.* [2000] notano che due elementi fondamentali di un generico modello di cooperazione tra agenti BDI sono il considerare il vantaggio complessivo che il gruppo trae dalle decisioni di un singolo agente e la modellazione ricorsiva, effettuata da ogni singolo agente, delle decisioni dei compagni. Se questi due elementi sono presenti, il comportamento del gruppo soddisfa proprietà quali collaboratività, comunicazione, capacità di evitare conflitti, ...

## 6.9. Applicazione pratica del paradigma BDI

Il problema che si incontra nell'impostazione BDI degli agenti è la ridotta applicabilità pratica, ovvero la costruzione in un sistema software funzionante; già si è detto che, in generale, gli agenti di tipo deliberativo hanno come svantaggio i tempi necessari per l'elaborazione delle soluzioni. Gli agenti BDI, essendo progettati utilizzando paradigmi di logica modale, richiedono rilevanti risorse computazionali e, ancor prima, rendono decisamente arduo il compito di convertire una struttura così complessa in un qualsivoglia linguaggio di programmazione standard. Per questo motivo sono stati creati linguaggi *ad hoc*, che si occupano del *parsing* (letteralmente, dell'analisi grammaticale) delle formule logiche e della loro risoluzione analitica. Tuttavia, anche con questi linguaggi, è molto complesso rappresentare situazioni ampie e realistiche nell'ambito del paradigma; inoltre, se gli agenti in gioco sono in numero elevato ed eterogenei, la realizzazione è quasi impossibile. Questo fa sì che, al momento, la tecnologia BDI sia soprattutto utilizzata per progetti teorici, in cui è utile modellare una determinata situazione utilizzando un linguaggio logico, piuttosto che simularla dinamicamente.

## 6.10. L'interazione tra gli agenti e le applicazioni

Qualunque sia il tipo di agenti utilizzato, i risultati più interessanti si ottengono con l'interazione nell'ambito dei sistemi multi-agente o MAS. Le principali aree di applicazione per questo tipo di tecnica sono elencate qui di seguito.

### 6.10.1. Scienze sociali

Nell'ambito delle scienze sociali gli agenti reali e le loro interazioni sono simulati in un ambiente artificiale ricreato al calcolatore con lo scopo di replicare i comportamenti che si osservano nella realtà, di cui gli agenti del modello ed i loro comportamenti sono una rappresentazione in scala. La teoria dei sistemi ha dimostrato che molti sistemi reali, in apparenza molto diversi, mostrano proprietà intrinseche e strutturali simili; questo permette di riutilizzare un modello ad agenti per più casi diversi, apportando modifiche minori. In particolare i sistemi multi agente possono essere applicati con ottimi risultati ai sistemi complessi, quelli per cui è difficile, se non impossibile, restringere la descrizione ad un numero limitato di parametri o variabili caratteristiche, senza perdere le proprietà funzionali a livello globale.

Un sistema complesso mostra solitamente proprietà emergenti che non sono direttamente anticipabili o identificabili attraverso la comprensione delle componenti. L'*emergenza* è probabilmente la proprietà più interessante di un sistema complesso ed è definita come un fenomeno di livello più elevato, che non può essere ricondotto al funzionamento delle parti. L'emergenza è il processo che consiste nel derivare alcune strutture e proprietà, nuove e coerenti, in un sistema complesso. I fenomeni emergenti sono causati dalle interazioni – non lineari e distribuite – tra gli elementi del sistema, con il trascorrere del tempo. Una caratteristica dei fenomeni emergenti è che essi possono essere osservati soltanto a livello macro, anche se sono generati da elementi a livello micro.

Per questo motivo i sistemi multi-agente possono replicare in modo significativo questi fenomeni, soprattutto nei casi in cui gli agenti coinvolti siano prevalentemente di tipo reattivo. Inoltre i sistemi sociali complessi sono fondamentalmente non deterministici: è impossibile prevederne precisamente il comportamento anche quando si conosca la funzione di ogni componente. Questo rende di fatto praticamente impossibile lo studio di tali sistemi attraverso metodologie matematiche e statistiche, a meno di non operare profonde semplificazioni.

Nelle scienze sociali i sistemi MAS dell'informatica e le specificazioni degli agenti – di tipo BDI, o semplificate, o ricostruite ad esempio con reti neurali – sono state in gran parte riproposte da zero, in modo parallelo a quello dell'informatica, dando vita al filone dei modelli basati su agenti (*Agent Based Models* o ABM) cui questo libro è dedicato. La presentazione dei MAS ha soprattutto lo scopo di proporre l'avvicinamento e il dialogo tra le due visioni della simulazione multi agente, vista dal lato delle componenti software (MAS) e del modello scomposto ad agenti (ABM).

### 6.10.2. Scienze ingegneristiche

Per le scienze ingegneristiche, lo scopo della applicazione di modelli ad agenti è quello di progettare e sviluppare sistemi in modo innovativo. In questo caso le simulazioni basate su agenti possono essere utilizzate per condurre analisi di tipo *what-if*, cioè per studiare cosa accadrebbe se, ad un sistema esistente, si apportassero determinate modifiche. In questo ambito la simulazione basata su agenti si pone come valida alternativa alla più rigida e deterministica simulazione di processo, o ad eventi discreti, come illustrato in Remondino [2003].

### 6.10.3. Mondi virtuali

I sistemi multi agente possono essere utilizzati per sviluppare mondi virtuali [Damer, 1998] all'interno dei quali anche gli esseri umani hanno la possibilità di interagire con il sistema, tramite degli *avatar*<sup>1</sup>, cioè dei corrispettivi informatici che eseguono, nel mondo simulato, i comandi ad essi impartiti dagli utilizzatori. Questo tipo di impiego è soprattutto utile in situazioni idealizzate in cui si desidera verificare l'efficacia di teorie, che normalmente nella realtà sarebbero inquinate da effetti esterni indesiderati.

---

<sup>1</sup>si veda il cap. 22