

UNIVERSITA' POLITECNICA DELLE MARCHE

Facoltà di Ingegneria

Corso di laurea in Informatica



# **Modellazione di stati mentali in architetture BDI per “Software Agents”**

Candidato:  
Matteo PIERSANTELLI

Relatore:  
Prof. Aldo Franco DRAGONI

Correlatore:  
Prof. Paolo PULITI

ANNO ACCADEMICO 2009/2010



## **ABSTRACT**

Dragoni ha sviluppato una nuova struttura logica per rappresentare il comportamento di un agente, soprattutto nel momento di interazione con altri agenti. Tuttavia, non sono mai stati effettuati degli studi per capire se tale struttura è rappresentata da una sintassi e una semantica valide. Quindi dopo uno studio della struttura e delle teorie che la sorreggono si è completata la semantica della struttura che presentava diverse incomprensioni e mancava di alcune definizioni. Inoltre è stata aggiunta una caratteristica che permette di avere un legame tra le diverse attitudini mentali. Infine sono state introdotte le linee guida per arrivare a studiare le dinamiche della struttura in presenza di atti comunicativi.



# INDICE

---

INDICE.....	5
1 INTRODUZIONE.....	7
1.1 Agenti e sistemi Multi-Agente.....	8
1.2 Agenti.....	13
1.3 Architettura BDI.....	16
1.4 Una nuova struttura logica.....	17
2 LA STRUTTURA DELLO STATO MENTALE.....	21
2.1 Struttura mentale.....	22
2.2 I fatti del mondo.....	27
3 ESTENSIONE DELLA SEMANTICA.....	33
3.1 Perché estendere la semantica?.....	34
3.1.1 L'indeterminismo.....	34
3.1.2 Perché abbiamo bisogno dell'indeterminismo?.....	38
3.2 Revisione della semantica.....	44
4 INSIEME DEI VINCOLI DI RAGIONAMENTO.....	51
4.1 Natura dei vincoli.....	52
4.2 Semantica.....	53
4.3 Conclusioni.....	55
5 COMUNICAZIONE TRA AGENTI.....	57
5.1 Introduzione.....	58
5.2 Agent Communication Language.....	58
5.3 Teoria degli speech act.....	63
5.4 Conclusioni.....	66

6	DINAMICA DELLA STRUTTURA MENTALE .....	67
6.1	Introduzione.....	68
6.2	Linee guida .....	69
	BIBLIOGRAFIA .....	73

# 1 INTRODUZIONE

---

## 1. INTRODUZIONE

### 1.1 Agenti e sistemi Multi-Agente

L'idea degli agenti non è recente, il genere umano ha sempre sognato di creare l'intelligenza artificiale e le forme di vita artificiali, questo con lo scopo di essere d'aiuto per il nostro business quotidiano. L'idea degli agenti infatti riguarda l'apprendimento e il supporto per l'utente.

Secondo Allan Kay l'idea dei software agents va attribuita a John McCarthy e Oliver G.Selfridge durante il loro lavoro nel 1950 al MIT.

*“They had in view a system that, when given a goal could carry out the details of the appropriate computer operations and could ask back for and receive advice, offered in human terms, when it was stuck. An agent would be a soft robot living and doing its business within the computer world.” [Kay, 1984]*

*“Loro avevano in mente un sistema che, dato un obiettivo potesse fornire dettagliatamente le operazioni computeristiche appropriate, chiedere e ricevere delle informazioni, e che fosse in grado di comunicare, in linguaggio comprensibile, un eventuale blocco. Un agente sarebbe una specie di robot che interagisce e compie un interscambio di informazioni all'interno del “computer world”.”*

Un agente che capace di dedurre piani di esecuzione da un dato obiettivo, al fine di effettuare adatte operazioni, e in grado di accettare istruzioni in linguaggio umano, conduce subito nel campo di ricerca dell'intelligenza artificiale (Artificial Intelligence), che aveva ed ha ancora oggi la principale influenza nello studio degli agenti. Jennings [Jennings et al., 1998] afferma che l'intelligenza artificiale è tutto ciò che riguarda la creazione di artefatti intelligenti e che, se questi artefatti possono percepire ed agire in determinati ambienti, allora possono essere considerati agenti. Alcuni gruppi specialistici all'interno dell'AI Community sono strettamente legati agli studi degli agenti. Sono chiamati Artificial Intelligence Planning (AIP) e Distributed Artificial Intelligence (DAI). La ricerca in queste aree fu



classificata da Nwana [Nwana, 1995] come il primo passo verso la storia dello sviluppo degli agenti. AIP è di solito associato con il sistema di pianificazione STRIPS (Stanford Research Institute Problem Solver) [Fikes and Nilsson, 1971], che usa una rappresentazione simbolica dell'ambiente dell'agente, delle sue capacità e dei suoi obiettivi per descrivere il sistema stesso. L'algoritmo di pianificazione impiega tecniche di ragionamento simboliche per creare uno schema. Quest'ultimo può essere eseguito dall'agente in modo da raggiungere i suoi obiettivi. Negli anni '80 fu chiaro che, sebbene l'algoritmico di pianificazione STRIPS e i suoi risultati fornissero prestazioni ragionevoli, essi non si adattavano alla perfezione ad esempi del mondo reale. Rodney Brooks è stato uno dei precursori nell'applicare il ragionamento simbolico dell'intelligenza artificiale ai fatti della vita reale. Basato sulla sua esperienza egli propose un'alternativa alla AI simbolica conosciuta anche come AI del comportamento, AI reattiva o AI collocata.

Distributed Artificial Intelligence può essere suddivisa in Distributed Problem Solving (DPS) e Multi Agent System (MAS). Il primo si concentra su come la conoscenza può essere distribuita fra differenti moduli che cooperando risolvono i problemi. Nei sistemi DPS le strategie di comunicazione e di interazione sono di solito predefinite e formano una parte integrante del sistema. Al contrario gli agenti in MAS sono risolutori di problemi autonomi, i quali sono debolmente accoppiati e mirano alla risoluzione di problemi che altrimenti sarebbero oltre le capacità di ogni singolo agente. Sebbene i MAS prendono origine dall'AI, questa parola è oggi spesso usata in un senso differente: per denotare sistemi che possono essere visti come l'insieme di diversi agenti.

Un secondo filone di sviluppo di un agente è stato identificato da Nwana [Nwana, 1995] nei primi anni 90 con la diversificazione in tipi di agenti. Questo era il tempo in cui il termine agente è stato preso da ricercatori di campi di studio differenti quali robotica, vita artificiale, distributed object computing, interazione uomo-computer ecc. Da allora la nozione, da sempre non ben definita di cosa è un agente, è diventata ancora più vaga e la categorizzazione dei tipi di agente è sempre più complessa. Nwana vede l'Attore, modello proposto da Agha e Hewit come uno dei

## 1. INTRODUZIONE

primi modelli per descrivere MAS. In questo modello l'Attore è un oggetto autonomo interattivo e continuamente in esecuzione. La necessità di queste tre proprietà è largamente accettata.

*“An actor is a computational agent which has an email address and a behaviour. Actors communicate by message passing and carry out their actions concurrently.”*[Nwana, 1995]

*“Un attore è un agente computazionale che ha un indirizzo e un comportamento. Gli attori comunicano tramite messaggi ed esercitando azioni con correntemente.”*

Innegabilmente altri campi di ricerca quali la programmazione orientata agli oggetti (OOP), sistemi concorrenti object-based, così come la progettazione di interfacce uomo-computer (HMI) contribuiscono significativamente alla corrente ricerca in agenti. Zambonelli e altri affermano in [Zambonelli et al., 2003] che per i ricercatori nel campo dell'OOP gli agenti sono spesso visti come oggetti. Sebbene è vero che oggi la maggior parte delle applicazioni basate su agenti sono scritte in ambienti OOP, gli agenti sono visti come un modo più astratto di sistemi altamente interattivi e distribuiti. Un approccio di sviluppo MAS – la metodologia GAIA – è presente in [Zambonelli et al., 2003]. E' basata sulla identificazione di organizzazioni e ruoli, che successivamente nello sviluppo vengono assegnati ad agenti concreti.

L'idea di utilizzare agenti come un astrazione e MAS direttamente creati dalla programmazione e dalle loro relazioni fu proposta da Shoham negli anni 90. La sua idea si manifestò nel paradigma AOP, come astrazione dell'OOP, che descrive e implementa le entità (agenti), le loro relazioni e le loro interazioni. Il suo lavoro risultò in uno dei primi linguaggi ad agenti chiamato AGENT0, nel quale un agente è essenzialmente descritto in termini di capacità, credenze iniziali ed obblighi, chiamate regole di impegno, che sono responsabili dell'azione che l'agente eseguirà. Da allora altri ricercatori inventarono linguaggi ad agenti per descrivere MAS [Zambonelli et al., 2003] [Fisher, 1994] [Mascardi et al., 2005].

Jennings [Jennings et al., 1998] comparò in modo critico le differenze tra AOP e OOP. Una delle più evidenti differenze è il livello di astrazione e composizione che gli agenti forniscono. Diversamente un oggetto, che incapsula metodi e attributi, che possono essere accessibili da altri oggetti, un agente aggiunge anche l'incapsulamento del controllo su azioni e stati. Questo permette il comportamento dinamico che è spesso associato agli agenti [Zambonelli et al., 2003].

La ricerca nel campo degli agenti oggi può essere vista come l'unione di differenti tipi di aree, non solo limitate all'informatica. Filosofia e psicologia, ad esempio, hanno avuto e avranno un'importante influenza sugli agenti di oggi. Il filosofo Dennett coniò il termine *sistemi intenzionali*, in cui denotò sistemi il cui comportamento delle entità può essere predetto grazie all'attribuzione di credenze, desideri e funzioni razionali. Una piccola lista di queste attitudini che emergono dagli stati intenzionali si trova in [Wooldridge and Jennings, 1995]. Tre di queste attitudini formano la base dell'architettura BDI di Rao e Georgeff [Rao and Georgeff, 1995], che probabilmente è una delle migliori architetture conosciute che caratterizzano gli stati mentali degli agenti.

L'aspetto psicologico degli agenti è più ovvio quando *assistenti esperti* entrano in gioco. In questi sistemi un'interfaccia utente dinamica segue ed assiste l'utente nel suo lavoro. I software agents che seguono l'utente di sistema nel lavoro giornaliero diventano sempre più importanti e complessi. Nicholas Negroponte descrive il suo modo di vedere gli agenti in [Negroponte, 1997]:

*The best metaphor [...] for a human-computer interface is that of a well-trained English butler. The "agent" answers the phone recognizes the callers, disturbs you when appropriate, and may even tell a white lie on your behalf."*

*"La migliore metafora per un'interfaccia uomo-computer è quella di un ben allenato maggiordomo inglese. L'agente risponde al telefono riconoscendo chi chiama, disturbandoti quando appropriato, e può sempre dire una bugia per conto tuo."*

## 1. INTRODUZIONE

Per essere capace di aiutare in modo utile e produttivo, l'agente deve possedere la conoscenza riguardo l'utente. Per questo motivo la conoscenza non può essere espressa staticamente per un programma specifico (come si fa tradizionalmente) ma necessita di adattarsi ai cambiamenti delle esperienze e preferenze dell'utente, imparare diventa quindi di estrema importanza per gli agenti che hanno il ruolo di *assistenti esperti*. Pattie Maes enfatizza che l'agente:

*“[...] does not act as an interface or layer between the user and the application. It rather behaves as a personal assistant which cooperates with the user on the task. The user is able to bypass the agent.”*[Maes, 1997]

*“non funge da interfaccia o strato tra l'utente e l'applicazione. Piuttosto si comporta come un assistente personale che coopera con l'utente sul lavoro. L'utente è capace di bypassare l'agente.”*

Questa cooperazione e l'apprendimento attirano ancora attenzione nel campo dell'AI. L'elaborazione del linguaggio naturale, il riconoscimento della parola [Ball et al., 1997], e l'apprendimento da dimostrazione [Smith et al., 1997] giocano un ruolo importante nello sviluppo di assistenti personali.

Un altro importante risultato che è stato già menzionato si presenta con la scoperta di MAS. Tradizionalmente, i MAS per la risoluzione di problemi distribuiti erano sistemi chiusi, in cui tutti gli agenti nel MAS sono conosciuti e cooperano tra di loro. Con il secondo filone di sviluppo di agenti, e specialmente con l'invenzione degli agenti mobili [Chess et al., 1994], che migrano da un server ad un altro attraverso la rete, in modo da completare un lavoro, la cooperazione degli agenti non può essere presupposta.

Questa corta introduzione della storia dello sviluppo di agenti serve a dare un'idea di come sono complessi e diversificati. Per riassumere: l'idea di agenti inizia nel campo dell'intelligenza artificiale ed è stata applicata in altre discipline nei primi anni 90. L'AI richiama la funzione di apprendimento, pianificazione e collaborazione di vari aspetti degli agenti

come mobilità, tolleranza agli errori, interazione con utenti ecc. e le introduce all'interno di un unico contesto.

## 1.2 Agenti

Sfortunatamente il campo di ricerca in agenti è vasto e la maggior parte delle definizioni che sono date agli agenti sembrano provenire dalla specifica applicazione che l'autore ha in mente. Nwana afferma che c'è:

*"[...] as much chance of agreeing on a consensus definition for the word 'agent' as AI researchers have of arriving at one for 'artificial intelligence' itself."*[Nwana, 1995]

*"tanta probabilità di accordarsi su una definizione per la parola 'agente' quanta ne hanno i ricercatori di AI di arrivare ad una definizione per 'intelligenza artificiale' stessa."*

Di fronte a questo, sembra essere obbligatorio partire da una definizione di agente che l'autore sostiene all'inizio di ogni pubblicazione. Questa sezione innanzitutto offre delle definizioni di agenti date da differenti autori, per poi descrivere le proprietà più comunemente associate agli agenti.

Benché Nwana ritiene una definizione consensuale di agente come infattibile e preferisce considerare il termine agente come un termine generico, egli dà la seguente descrizione di agente:

*"[An agent is] a component of software or hardware which is capable of acting exactly in order to accomplish tasks on behalf of its user."*[Nwana, 1995]

*"[Un agente è] un componente software o hardware che è capace di agire esattamente in modo da compiere task per conto dell'utente."*

Questa nozione di agente è probabilmente una delle più appropriate e sembra essere molto quotata. Secondo il Webster Dictionary [Webster, 1913] la parola deriva dal greco *agere*, agire. Definisce l'agente come:

## 1. INTRODUZIONE

1. Colui che esercita il potere o ha il potere di agire; un attore.
2. Colui che agisce per, o al posto di, in nome e per conto di qualcun altro; un sostituto, un delegato; elemento, fattore.
3. Un potere attivo o causa; che ha la capacità di produrre un effetto; sia esso agente fisico, chimico o medico.

L'abilità di agire è la proprietà principale che è associata all'agente. Nella precedente definizione il verbo agire è ancora molto generale. Shoham dà una definizione che è accettata nella AI Community. Questa definizione è più concreta nel descrivere l'abilità di agire dell'agente.

*“Most often, when people in AI use the term 'agent,' they refer to an entity that functions continuously and autonomously in an environment in which other processes take place and other agents exist.”*[Shoham, 1997]

*“Molto spesso quando le persone in AI usano il termine 'agente', si riferiscono ad un'entità che funziona continuamente e autonomamente in un ambiente dove prendono posto altri processi e esistono altri agenti.”*

Questo dà risalto alla continuità e all'autonomia dell'agente e denota esplicitamente l'importanza dell'ambiente in cui l'agente è attivo. Ugualmente importante è l'impatto dell'ambiente nella definizione data da Franklin e Graesse:

*“An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.”*[Franklin and Graesser, 1996]

*“Un agente autonomo è un sistema situato dentro e in una parte dell'ambiente che percepisce l'ambiente e in esso agisce, col passare del tempo, insegue i suoi compiti in modo da effettuare una percezione futura.”*

Qui è messa in risalto l'autonomia e la relazione tra gli agenti e il loro ambiente. Da questa relazione nasce una parte cruciale della definizione di Franklin e Graesser che afferma che i sistemi sono agenti o meno in relazione agli ambienti. Questo fa riflettere sul fatto che se l'ambiente

dell'agente è cambiato, l'agente può perdere il suo stato in riferimento al suo ambiente. Una definizione simile è data da Wooldridge & Jennings.

*“An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.” [Wooldridge, 1995]*

*“Un agente è un sistema informatico che si trova in un ambiente, ed è capace di azioni autonome nel suo ambiente al fine di raggiungere gli obiettivi programmati.”*

Questa definizione si focalizza sulla posizione dell'agente nel suo ambiente, in autonomia e orientato agli obiettivi. La lista delle definizioni può andare oltre, ma questa menzionata è sufficiente a riportare le seguenti proprietà che un agente può esibire. La seguente lista di proprietà è adottata da Bradshaw:

**Reattività:** l'abilità di percepire e agire selettivamente.

**Autonomia:** orientato agli obiettivi, è dinamico e ha un comportamento autonomo.

**Comportamento collaborativo:** può lavorare insieme ad altri agenti per perseguire uno scopo comune.

**Abilità comunicativa:** l'abilità a comunicare con persone ed altri agenti con un linguaggio più simile all'atto comunicativo umano che al tipico livello simbolico di programmazione dei protocolli.

**Capacità inferenziale:** può agire su compiti astratti usando specificatamente la conoscenza degli obiettivi generali e dei metodi prefissati per raggiungere la flessibilità; va oltre l'informazione data e può avere espliciti modelli dell'utente, della situazione e/o dell'agente.

**Continuità temporale:** persistenza di identità e stati su un lungo periodo di tempo.

**Personalità:** la capacità di manifestare gli attributi di un carattere attendibile come l'emozione.

**Capacità di adattamento:** essere capaci di apprendere e migliorare con l'esperienza.

## 1. INTRODUZIONE

**Mobilità:** essere capaci di spostarsi in un modo auto didatta da una piattaforma ad un'altra,

Queste proprietà sono state anche identificate da Franklin e Graesser e si riflettono nella loro definizione. Per loro le proprietà di reattività, autonomia, orientamento agli obiettivi e continuità temporale sono essenziali per gli agenti, visto che le altre sono utilizzate per classificare la tipologia di agente.

### 1.3 Architettura BDI

Il filosofo Daniel Dennett ha coniato il termine *sistemi intenzionali* per descrivere quelle entità il cui comportamento può essere predetto dal metodo di attribuzione delle credenze, dei desideri e della perspicacia razionale [Dennett, 1987]. Grazie anche agli scritti di [McCarthy, 1978], [Shoham, 1990] e [Shardlow, 1990] possiamo dire che un agente è convenientemente descritto da atteggiamenti intenzionali. Esistono due importanti categorie di attitudini per rappresentare un agente che sono le attitudini di informazioni e le pro-attitudini. La prima categoria racchiude solamente la conoscenza e la credenza, queste attitudini sono legate all'informazione che un agente ha riguardo al mondo in cui si trova. Invece la seconda categoria è molto più grande e prende in considerazione i desideri, le intenzioni, gli obblighi, gli impegni, le scelte ecc., queste pro-attitudini sono quelle che guidano in qualche modo le azioni di un agente. Non c'è ancora tutt'oggi un chiaro consenso né in Intelligenza Artificiale né nella comunità filosofica riguardo ad una precisa combinazione di attitudini di informazione e pro-attitudini più adatta a caratterizzare un agente razionale. Tuttavia è stata riconosciuto, sia nella letteratura filosofica che in Intelligenza Artificiale, il ruolo stabilito dalle attitudini come credenze (Beliefs), desideri (Desires o Goal) e intenzioni (Intensions) nella progettazione di agenti razionali [Bratman, 1987] [Bratman et al., 1988] [Georgeff and Ingrand, 1989]. Alcuni degli aspetti filosofici delle teorie di



Bratman sono stati formalizzati da Cohen e Levesque [Cohen e Levesque, 1990]. Partendo dai lavori di Bratman, Cohen e Levesque, è stata sviluppata, da Rao e Georgeff, una struttura logica per la teoria di agenti basata su tre modalità primitive: credenze (B), desideri (D) e intenzioni (I) [Rao and Georgeff, 1991a], [Rao and Georgeff, 1991b], [Rao and Georgeff, 1993]. In questa architettura le credenze rappresentano la conoscenza del mondo in cui si trova l'agente; i desideri rappresentano gli obiettivi che un agente deve compiere; le intenzioni rappresentano le azioni attualmente condotte per raggiungere i desideri dell'agente.

La nozione di un agente intelligente come un'entità che appare essere l'argomento di attitudini mentali come credenze, desideri e intenzioni è ben nota e riconosciuta da molti ricercatori. Dietro la definizione delle varie logiche BDI, molti linguaggi e ambienti integrati per la programmazione di agenti BDI sono stati proposti sin dagli anni 90. Troviamo agenti software programmati sulla base di questa architettura in campo industriale come il trattamento dei malfunzionamenti dello space shuttle, sistemi di controllo del traffico aereo, UAV (veicoli aerei senza equipaggio), agenti animati in ambienti virtuali ecc.

Un'introduzione ai diversi ambienti integrati per la programmazione di agenti BDI e il loro utilizzo in campo industriale può iniziare da [Mascardi and Eiter, 2002] [Mascardi et al, 2005].

## **1.4 Una nuova struttura logica**

Nei sistemi multi-agente, se gli agenti non sono progettati con conoscenza precompilata delle credenze, intenzioni, abilità e prospettive di altri agenti, hanno bisogno di scambiare informazioni in modo da cooperare e coordinare le loro attività. Comunque, in domini di applicazioni reali, la comunicazione potrebbe essere una risorsa limitata (banda limitata, basso rapporto segnale/rumore ecc.). In alcuni casi, per gli agenti, risulta essere molto importante, quando decidono di mandare un messaggio, considerare i

## 1. INTRODUZIONE

benefici attesi in rapporto al costo di comunicazione [Gmytrasiewicz et al., 1991].

In modo da valutare l'utilità della comunicazione, gli agenti devono essere capaci di prevedere gli effetti delle loro dichiarazioni sullo stato mentale del destinatario. Per questo scopo, uno speaker può provare a riconoscere e modellare lo stato mentale dell'ascoltatore e poi predire l'impatto del suo messaggio sulle basi di questo modello. Il riconoscimento degli stati mentali diventa allora molto importante per agenti che effettuano comunicazioni efficaci. Questo compito può essere portato a termine dalle osservazione dei comportamenti degli altri agenti [Rao and Murray, 1994] e dal contenuto delle loro dichiarazioni [Dragoni and Puliti, 1994]. Assumendo che la comunicazione è il principale modo per scambiare porzioni di stati mentali, ci si deve focalizzare su come riconoscere lo stato mentale dello speaker sulla base delle sue dichiarazioni.

Gli agenti BDI sono supposti avere uno stato mentale, il quale contiene credenze (beliefs), desideri (desires) e intenzioni (intention), riguardo l'ambiente in cui si trovano e riguardo le credenze, i desideri e le intenzioni degli altri agenti. Il comportamento di un agente è strettamente dipendente dal suo stato mentale e la comunicazione è generalmente utilizzata per influenzare il comportamento di altri agenti. La comunicazione è supposta essere intenzionale, cioè attivata dal ragionamento dello speaker riguardo le sue credenze, i suoi desideri e le sue intenzioni. In altre parole, è generalmente possibile considerare l'espressione come la conseguenza di un particolare stato mentale dello speaker che provoca il desiderio di influenzare lo stato mentale dell'ascoltatore. Se le dichiarazioni sono il risultato di un ragionamento basato sulle condizioni mentali, sembra naturale supporre che l'ascoltatore può utilizzare la comunicazione ricevuta come base di un ragionamento a ritroso, al fine di riconoscere l'ipotetico stato mentale dello speaker che ha originato la comunicazione. Il ragionamento abduttivo può essere di aiuto per questo scopo. In particolare, l'ascoltatore può utilizzare l'abduzione come ragionamento a ritroso in base al tipo e al contenuto della comunicazione ricevuta.

Sebbene agenti BDI sono stati profondamente studiati sia da una prospettiva teorica sia da una pragmatica, meno attenzione è stata data alla struttura ricorsiva degli stati mentali, i quali giocano un ruolo essenziale quando modelliamo una interazione di alto livello tra agenti intelligenti. Lo studio di Dragoni [Dragoni, 2009] prova a catturare questa proprietà introducendo un approccio multi contestuale per la rappresentazione degli stati mentali. Una semantica per il formalismo multi contestuale è fornita basandosi sulla definizione di “struttura mentale”, la quale è un traliccio gerarchico di moduli triangolari  $\langle x, B, D \rangle$ . La componente  $x$  rappresenta lo stato mentale dell’agente  $x$  nell’insieme, mentre le componenti  $B$  e  $D$  rappresentano rispettivamente le credenze e i desideri dell’agente  $x$ . Se altre attitudini mentali, come intenzioni e impegni, sono considerate come primitive, allora esse possono essere inglobate nel modulo base, altrimenti devono essere rappresentate in termini di credenze e desideri.

Il lavoro svolto in questa tesi parte dallo studio effettuato da Dragoni [Dragoni, 2009]. Dragoni ha sviluppato una nuova struttura logica per rappresentare il comportamento di un agente, soprattutto nel momento di interazione con altri agenti o con una persona umana. La struttura introdotta da Dragoni è completa di sintassi e semantica che sostengono la teoria. Lo scopo principale è quello di riuscire a capire come cambia lo stato mentale di un agente nel momento in cui riceve una comunicazione dall’esterno, e capire da cosa scaturisce il fatto di mandare questa comunicazione verso l’esterno. Quindi il lavoro si basa sull’atto comunicativo, o speech act, tra diversi agenti, sulle precondizioni che portano alla generazione dell’atto comunicativo, sul contenuto del messaggio dell’atto comunicativo e sulle post condizioni generate dall’atto comunicativo, sia da parte del mittente sia del destinatario. La struttura introdotta da Dragoni è studiata per arrivare a soddisfare quanto detto. Tuttavia, non sono mai stati effettuati degli studi per capire se tale struttura è rappresentata da una sintassi e una semantica valide. Quindi in un primo momento ho dovuto effettuare uno studio della struttura e delle teorie che la sorreggono. Successivamente ho completato la semantica della struttura che presentava diverse incompiutezze e mancava di alcune definizioni. Inoltre ho aggiunto una caratteristica che permette di

## 1. INTRODUZIONE

avere un legame tra le diverse attitudini. Infine ho introdotto le linee guida per arrivare a studiare le dinamiche della struttura in presenza di atti comunicativi.

Nel secondo capitolo troviamo la parte che concerne la struttura mentale, introducendo una definizione e spiegando il perché utilizzarla. Il terzo capitolo riguarda la semantica della struttura e del sistema multi contestuale. Più precisamente la semantica viene in parte estesa e in parte corretta rispetto al lavoro di Dragoni. Il quarto capitolo riguarda i vincoli di ragionamento che permettono di creare un legame tra le diverse attitudini mentali. Nel quinto capitolo introduciamo la comunicazione tra agenti, spiegando il campo di ricerca attuale in questo indirizzo, non essendo l'obiettivo della presente tesi quello di sviluppare un protocollo per la comunicazione tra agenti basati su questa struttura innovativa. Infine il sesto capitolo ha lo scopo di indirizzare una fase di ricerca successiva per lo studio della dinamica della struttura mentale, nel momento in cui un agente interagisce tramite atti comunicativi in un gruppo di agenti o con esseri umani.

## 2 LA STRUTTURA DELLO STATO MENTALE

---

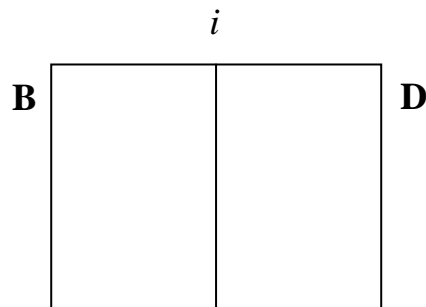
### 2.1 Struttura mentale

La nozione di una rappresentazione mentale è un concetto base dell'Intelligenza Artificiale, secondo cui gli stati cognitivi e i processi sono costituiti dalle avvenute trasformazioni e immagazzinamento delle strutture di supporto alle informazioni.

La rappresentazione mentale dipende dalla esistenza di attitudini proposizionali e dalla determinazione dei loro contenuti come proprietà fenomenali di pensiero e di percezione di esperienza.

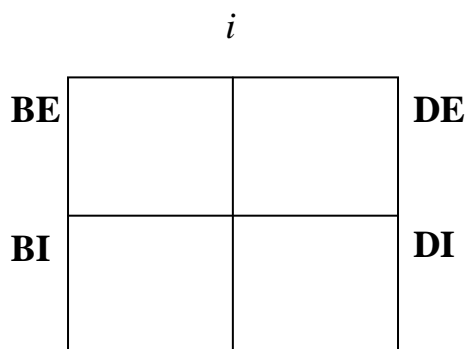
La maggior parte delle ricerche effettuate in questo campo hanno portato all'affermazione dell'architettura introdotta da Georgeff e Rao, una architettura che si basa sulle attitudini proposizionali Beliefs-Desires-Intention (BDI). Al fine di semplificare il più possibile ogni rappresentazione in questa struttura introdotta da Dragoni, vengo prese in considerazione solamente due attitudini proposizionali: Beliefs e Desires. Questo non vuol dire che solamente queste due attitudini riescano a descrivere e prevedere il comportamento di un agente, ma sicuramente riusciremo in un secondo momento ad introdurre le attitudini mancanti o a derivare altre attitudini principali, come la conoscenza (knowledge) e le intenzioni (intention) [Rao and Georgeff, 1995].

Quindi partiamo con la rappresentazione della struttura mentale del nostro agente intelligente, prendendo in considerazione la situazione di un singolo agente. Abbiamo appena detto che ci baseremo su due attitudini proposizionali, Beliefs e Desires, quindi possiamo inizialmente rappresentare la struttura mentale di un agente *i* come un **mondo** (Figura 2.1), dove l'agente ha delle credenze e dei desideri riguardo al mondo in cui si trova, inoltre è capace di ragionare su queste due attitudini. Il mondo quindi è suddiviso in due **continenti**, uno relativo ai Beliefs (B) e un altro relativo ai Desires (D).



**Figura 2.1**

Questo può essere tradotto come: l'agente  $i$  crede un certo insieme di argomenti **B** del mondo e desidera un insieme di argomenti **D** del mondo stesso. A questo punto possiamo introdurre una suddivisione all'interno di ciascuna di queste due aree, perché così come si ha una rappresentazione del mondo esterno si ha anche una rappresentazione di se stessi, cioè del mondo interno. Quindi l'agente ha delle credenze che riguardano il mondo esterno e delle credenze che riguardano il proprio stato mentale, inoltre l'agente ha dei desideri che riguardano il mondo esterno e dei desideri che riguardano il suo stato mentale. Più precisamente avremo una rappresentazione esterna (**E**) dell'agente e una rappresentazione interna (**I**) dell'agente, e questo si traduce in un mondo suddiviso in quattro continenti (Figura 2.2).



**Figura 2.2**

Normalmente avremo che  $\mathbf{BE} \neq \mathbf{DE}$ ; infatti l'uguaglianza  $\mathbf{BE} = \mathbf{DE}$  significa che l'agente  $i$  è completamente soddisfatto del mondo esterno dove lui è ambientato. Lo stesso discorso può essere fatto per **BI** e **DI**.

Questa nuova suddivisione ci porta ad effettuare un ragionamento sulla nostra rappresentazione interna dello stato mentale. Perché introdurre una vista interna equivale ad avere uno specchio del nostro stato mentale,

## 2. LA STRUTTURA DELLO STATO MENTALE

ciò ci porta ad avere sia per i Beliefs sia per i Desires un ulteriore mondo dello stato mentale. (Figura 2.3).

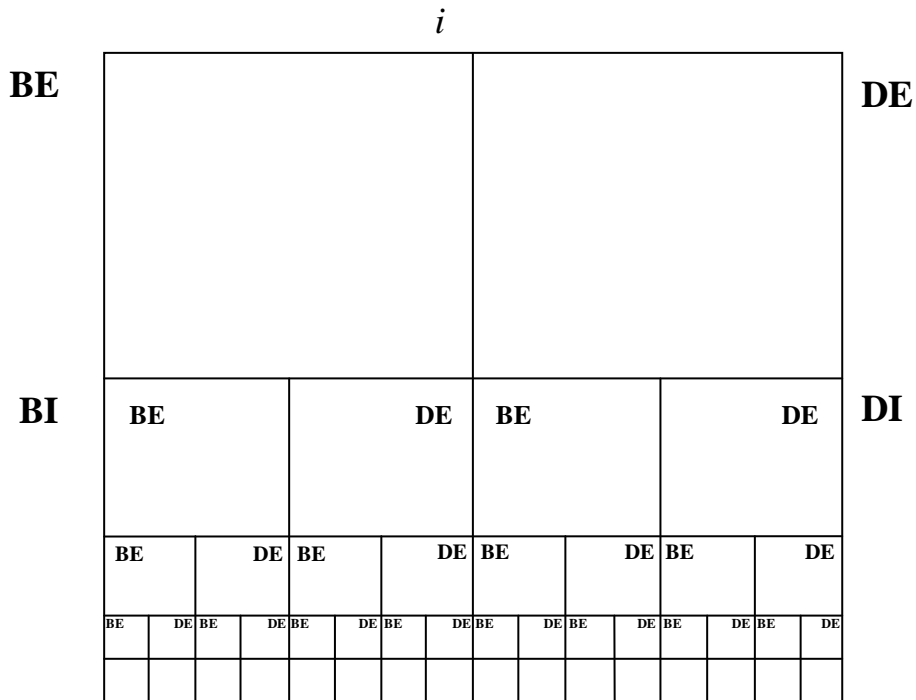


Figura 2.3

L'intuizione è che un contesto di credenza o di desiderio formalizza l'immagine mentale che l'agente ha di se stesso, o l'immagine mentale che ha dell'immagine mentale di se stesso, o ... Uno o più livelli dell'operatore Belief o dell'operatore Desire corrispondono a uno o più livelli della struttura dell'immagine mentale. Naturalmente risulta intuitivo che con questa introduzione lo stato mentale diventa ricorsivo, quindi le attitudini mentali primitive possono avere come argomento altre attitudini mentali. La

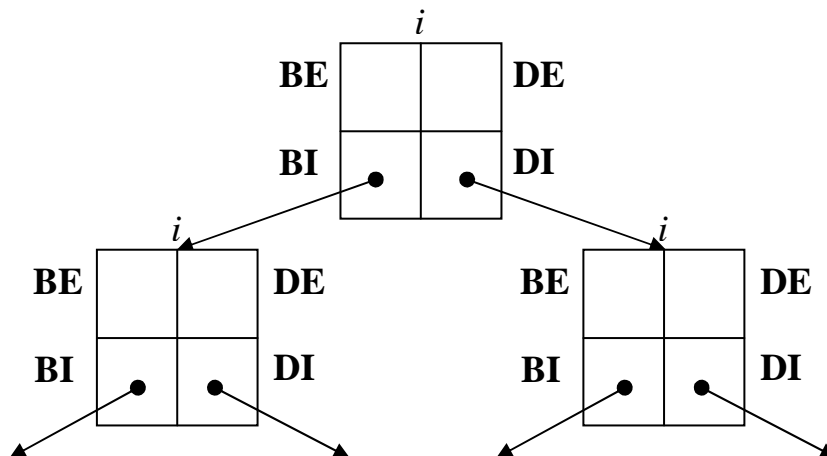


Figura 2.4



ricorsione è teoricamente infinita, anche se risulta essere sufficiente arrivare al terzo o quarto livello per rappresentare la maggior parte delle dinamiche mentali umane [Gmytrasiewicz et al., 1991] [Gmytrasiewicz and Durfee, 1995].

Quindi con questa rappresentazione troveremo nei riquadri **BI** e **DI** sempre una ricorsione della struttura mentale ad un livello inferiore (o interiore), mentre nei riquadri **BE** e **DE** troveremo i fatti del mondo che noi crediamo e desideriamo a quel livello. Il tutto può essere rappresentato con un albero (Figura 2.4).

Un caso d'uso, che risulta essere importante negli agenti software e nel campo dell'Intelligenza Artificiale, è quello dei sistemi multi agente in cui agenti intelligenti, che hanno differenti conoscenze e capacità, devono lavorare insieme per arrivare ad uno scopo, e per far questo hanno bisogno di comunicare nella maniera più appropriata le giuste informazioni per coordinare le loro azioni.

I vantaggi di una struttura di questo tipo si evidenziano nel momento in cui un agente software autonomo deve coordinarsi e comunicare in modo intelligente con altri agenti. Per far questo l'agente ha bisogno di ricrearsi la struttura mentale dell'agente con cui sta interagendo per capire come creare un messaggio capace di avere un impatto nel suo stato mentale. Quindi la nostra struttura mentale di un agente  $i$ , che deve comunicare con un altro agente  $j$ , non farà altro che ricrearsi la struttura mentale di questo secondo agente così come possiamo vedere nella Figura 2.5, dove, nella rappresentazione, ci fermiamo al terzo livello.

Grazie ad una struttura di questo genere l'agente dovrà controllare lo stato dei Beliefs e dei Desires dell'agente con cui vuole interagire per capire le precondizioni per un atto comunicativo, capire cosa deve contenere l'atto comunicativo e capire le post condizioni dell'atto comunicativo. Cioè risulta importante, per un agente, avere una riproduzione della struttura mentale di un altro agente il più veritiera possibile, in modo da scegliere nella maniera più corretta una comunicazione intelligente capace di modificare la struttura mentale del ricevente e modellando in anticipo la struttura mentale di

## 2. LA STRUTTURA DELLO STATO MENTALE

quest'ultimo in base all'effetto dell'atto comunicativo. Questo non vuol dire che l'agente avrà sicuramente una riproduzione fedele della struttura mentale dell'ascoltatore, ma sicuramente in caso di incomprendimento l'ascoltatore, a sua volta, effettuerà un atto comunicativo per cambiare la struttura mentale dell'altro agente. Come possiamo intuire, la comunicazione può andare avanti fino a che le rappresentazioni delle strutture mentali degli agenti non sono uguali.

A questo punto possiamo dare una definizione di struttura mentale.

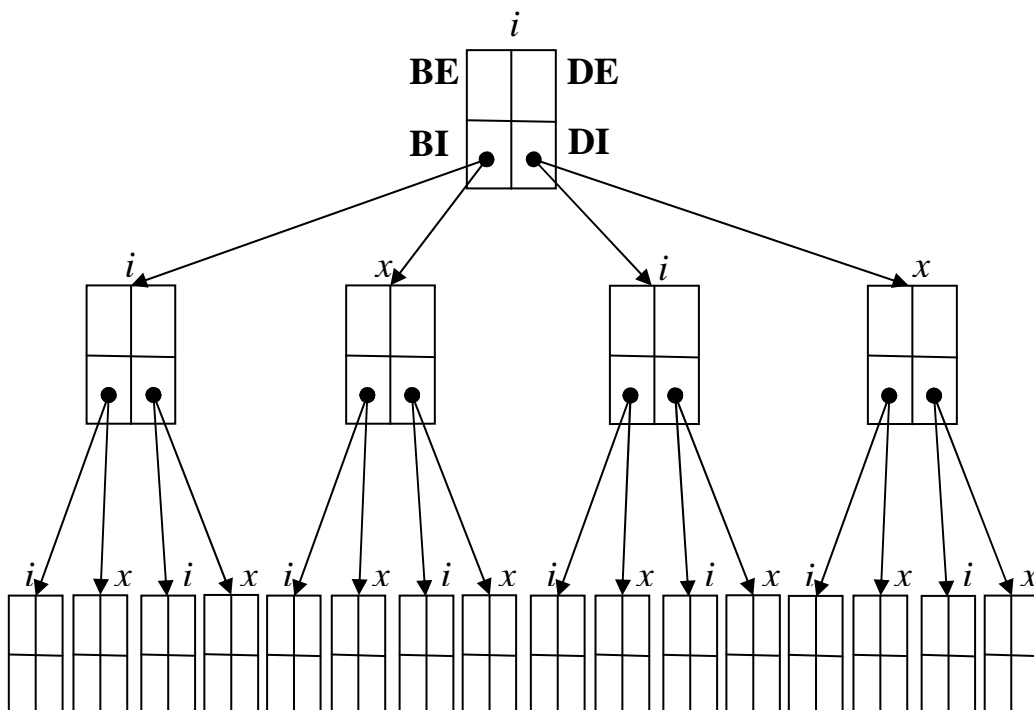


Figura 2.5

**Definizione 1.** La **struttura mentale** di un agente intelligente, definita con il simbolo  $\sum_{i,G}$  dove  $i$  è l'agente e  $G$  è il gruppo a cui appartiene, è un albero ricorsivo potenzialmente infinito, dove ogni nodo è una rappresentazione del mondo, suddiviso in mondo esterno, che raccoglie i fatti, e in mondo interno, che riflette (o rispecchia) la sua stessa rappresentazione più quella di tutti gli agenti appartenenti al gruppo  $G$ . Ogni mondo ha una suddivisione in due continenti, che rappresentano le attitudini mentali Beliefs e Desires.

Quindi diciamo che la struttura non è altro che uno scheletro dello stato mentale, che dovrà essere popolato con la conoscenza dell'agente.

## 2.2 I fatti del mondo

Abbiamo visto fino ad ora come deve essere creata la struttura mentale di un agente software, ma non abbiamo visto il contenuto di una struttura di questo tipo. Naturalmente, risulta essere la parte più importante capire come possiamo rappresentare i fatti del mondo. Come abbiamo detto i fatti verranno sempre rappresentati sotto forma di Beliefs e di Desires, inoltre dato il livello della struttura mentale, avremo anche Beliefs e Desires come argomenti di altrettanti Beliefs e Desires.

La rappresentazione del mondo esterno dove essere formata da un insieme  $\varphi$  che include tutte le proposizioni primitive, o fatti del mondo, sotto forma di linguaggio proposizionale. L'insieme  $\varphi$ , quindi, rappresenta tutti gli argomenti di cui è a conoscenza l'agente software sia che essi siano veri sia che essi siano falsi. Questo insieme verrà chiamato d'ora in avanti linguaggio esterno,  $EL_j$ , la cui definizione la troviamo in [Dragoni, 2009]. Prendendo tutti gli elementi dell'insieme  $EL_j$  possiamo creare una tabella di verità che dà come risultato una combinazione di 0 ed 1. Questa combinazione di 0 ed 1 viene chiamata funzione di verità che è una rappresentazione dei fatti del mondo. Successivamente parleremo di funzione di verità o di interpretazione allo stesso modo, poiché di quest'ultima ne è stata data la definizione in [Dragoni, 2009] Definizione 6. E' importante spiegare che i fatti del mondo di cui l'agente è a conoscenza, cioè che appartengono all'insieme  $EL_j$ , non saranno inclusi in tutti i livelli della struttura, ma ogni livello avrà una rappresentazione solo di una parte dei fatti, differenziando anche da quello che crede e da quello che desidera. Quindi ogni contesto potrà contenere differenti credenze del mondo e differenti desideri del mondo. Da quanto appena detto ne deriva che

## 2. LA STRUTTURA DELLO STATO MENTALE

l'interpretazione dei fatti dovrà essere effettuata in ogni casella **BE** o **DE** di ogni livello, cioè in ogni contesto troviamo una funzione di verità che da una rappresentazione dei fatti del mondo in quel contesto. Analizziamo ora i diversi tipi di rappresentazioni (Figura 2.6) che possiamo trovare all'interno di ogni contesto.

Partiamo prima dai due casi estremi che si riferiscono ad una rappresentazione con tutti 1 o con tutti 0. Nel primo caso, dove si hanno tutti 1, ci troviamo nella condizione in cui tutto è possibile, quindi l'agente non conosce niente, poiché l'agente crede o desidera tutto il possibile. Questa situazione è facilmente spiegabile prendendo come esempio un agente

	Una colonna	Più colonne
Un solo 1	1     0	3     0 1
	0	0 0
	1	0 0
	0	1 0
	0	0 0
Più 1	2     1	4     0 1
	1	1 1
	0	0 0
	1	1 0
	0	1 0
	0	0 1

**Figura 2.6**

intelligente che è stato appena creato e che ha l'insieme  $\varnothing$  non vuoto, ma non sa se gli elementi appartenenti a questo insieme sono veri o falsi, quindi la funzione di verità prende tutte le combinazioni possibili. Da questo esempio, possiamo anche capire che il fatto di aumentare la conoscenza dell'agente intelligente, si traduce nell'aggiunta di 0 al posto di 1 per quelle combinazioni dei fatti che non sono più possibili, o ancora meglio, non più veri.

Invece nel secondo caso, dove si hanno tutti 0, ci troviamo nella condizione in cui nessuna rappresentazione risulta possibile, quindi l'agente è in una situazione di contraddizione totale o inconsistenza. Cioè, l'agente intelligente con l'aumentare la sua conoscenza è arrivato ad annullare tutte le combinazioni possibili dei fatti. Naturalmente, non si vorrebbe mai

arrivare ad una situazione del genere perché non permette di effettuare nessun ragionamento, questo è dovuto dal fatto che ormai si hanno tutti 0 nella rappresentazione.

Passiamo ora a descrivere i quattro stati presenti nella Figura 2.6.

Nel riquadro 1 troviamo una rappresentazione di tutti 0 e di un solo 1, questo significa che l'agente in quel contesto, crede o desidera una sola combinazione o rappresentazione possibile (poiché in questo caso coincidono). Risulta abbastanza intuitivo capire il vantaggio di avere una rappresentazione di questo tipo in un contesto, poiché permette all'agente di avere una rappresentazione certa dei fatti presenti in quell'unico mondo possibile, cioè di non avere incertezze.

Il caso immediatamente successivo a quello appena visto è rappresentato nel riquadro 2 della Figura 2.6, in cui nella formula dobbiamo rappresentare più 1 contemporaneamente, cioè più combinazioni possibili. Il fatto di avere più combinazioni possibili significa che l'agente non è sicuro oppure è incerto riguardo ad alcuni fatti del mondo che sono rappresentati nel contesto che racchiude questa formula. Questo perché avendo più combinazioni possibili, al loro interno si avranno sicuramente alcuni fatti che sono uguali mentre altri si contraddicono. Con questa rappresentazione si introduce l'operatore *or* tra le diverse combinazioni possibili.

Fino ad ora abbiamo visto la possibilità di rappresentare o una sola combinazione possibile, fatti certi, o più combinazioni possibili, incertezza; ora introduciamo una situazione al quanto strana, in un primo istante, quella rappresentata dal riquadro 3 della Figura 2.6. Troviamo in questa rappresentazione più funzioni di verità all'interno di un continente della struttura mentale dell'agente intelligente. Subito vediamo che una rappresentazione di questo tipo indica che un agente crede o desidera, in modo certo, i fatti appartenenti ad una sola combinazione possibile di una funzione di verità e allo stesso istante crede o desidera in modo certo i fatti appartenenti ad un'altra combinazione possibile, differente dalla precedente, di un'altra funzione di verità.

## 2. LA STRUTTURA DELLO STATO MENTALE

Cioè viene introdotto l'operatore *or* tra insiemi di funzioni di verità e non tra i fatti rappresentati dalle combinazioni possibili. Quindi implica l'introduzione dell'operatore *or* tra le attitudini mentali, tra i Believes o tra i Desires, nella rappresentazione contestuale e multi-contestuale degli stati mentali, che vedremo nel capitolo successivo. Ogni casella della struttura mentale dell'agente potrà quindi avere o una colonna o un vettore di colonne per rappresentare il suo stato mentale. Con la presenza di più colonne ed un solo 1 su queste, significa che l'agente non conosce la verità riguardo ad un unico fatto del mondo, ma è sicuro di potersi trovare in un secondo momento o in una combinazione possibile o in un'altra combinazione possibile.

Il ragionamento può essere facilmente esteso al caso in cui si hanno più colonne con diversi 1 presenti, come nel riquadro 4 della Figura 2.6, dove abbiamo che l'agente si trova nella condizione di avere presenti nel suo stato mentale diversi fatti del mondo che non conosce, ma è certo che si potrà trovare in un secondo momento o in un insieme di combinazioni possibili o in un altro, cioè potrà avere o una funzione di verità o un'altra.

Il fatto di specificare che, solamente in un secondo momento l'agente potrà definire quale funzione di verità sia corretta, significa che l'agente grazie alla conoscenza potrà "filtrare" il suo stato mentale. Quindi, lo scopo dell'agente è sempre quello di arrivare ad una rappresentazione con una funzione di verità del tipo nel riquadro 1 della Figura 2.6, per far questo l'agente dovrà aumentare la sua conoscenza tramite l'interazione con il mondo esterno o con attori esterni, che possono essere persone umane o altri agenti software, partendo comunque da uno stato di conoscenza che è rappresentato da più 1 in un'unica colonna, o anche tutti 1, a rappresentare incertezza o non conoscenza, oppure con più colonne, a rappresentare situazioni certe in cui arriverà nel momento di apprendere conoscenza.

Le prime due rappresentazioni precedentemente discusse, le possiamo trovare all'interno dell'articolo [Dragoni, 2009], insieme alle loro definizioni e proprietà. Ora introduciamo le definizioni e le proprietà per le altre due rappresentazioni. Come abbiamo detto in precedenza il caso 4

estende solamente le caratteristiche del caso 3, ereditando le stesse definizioni e le stesse proprietà. Ma prima diamo una spiegazione sul perché vogliamo avere anche queste due interpretazioni.

## 2. LA STRUTTURA DELLO STATO MENTALE



## 3 ESTENSIONE DELLA SEMANTICA

---

## 3.1 Perché estendere la semantica?

### 3.1.1 L'indeterminismo.

La formula del tipo:

$$i: iB(a) \vee iB(\neg a)$$

appartiene al formalismo introdotto da [Dragoni, 2009], ma subito sembra produrre un significato un po' strano, poiché sta ad indicare che l'agente  $i$  crede  $a$  o l'agente  $i$  crede  $\neg a$ . In un primo istante potremmo semplicemente dire che l'agente  $i$  non è perfettamente consapevole dell'argomento  $a$ . Ma la corretta attitudine mentale che racchiude la primitiva di incertezza dell'agente  $i$  sull'argomento  $a$  è data dalla formula:

$$i: iB(a \vee \neg a)$$

Questa formula esprime il significato che l'agente  $i$  crede  $a$  o  $\neg a$ , quindi la credenza dell'agente  $i$  è perfettamente determinata, invece è il suo argomento ad essere indeterminato. Questa frase racchiude il concetto di tutto l'argomento che stiamo trattando e vedremo più tardi il perché.

Le due formule appena viste hanno un significato diverso poiché nella prima l'agente  $i$  è incerto, cioè non è perfettamente consapevole delle sue credenze (potenziale forma di indeterminismo), mentre la seconda formula indica che l'agente  $i$  è perfettamente consapevole di avere una credenza indefinita.

A questo punto viene dimostrato tramite il [Dragoni, 2009]Teorema 1 che applicando una semantica del riquadro 1 della Figura 2.6, cioè avendo un'unica interpretazione degli elementi appartenenti ad  $EL_i$ , si va incontro ad una equivalenza del tipo:

$$i: iB(a) \vee iB(\neg a) \equiv i: iB(a \vee \neg a)$$

Successivamente viene cambiata la semantica introducendo un approccio simile a quello dei mondi possibili, cioè come quello

rappresentato nel riquadro 2 della Figura 2.6, in modo da evitare questa uguaglianza. Il risultato è (Teorema 1b):

$$i: iB(a) \vee iB(\neg a) \rightarrow i: iB(a \vee \neg a)$$

Lasciamo la dimostrazione, ma per i lettori più curiosi è possibile vederla nell'articolo [Dragoni, 2009].

Il nostro discorso si focalizza sulla prima parte di questa formula cioè:

$$i: iB(a) \vee iB(\neg a)$$

La formula appena scritta, come precedentemente affermato, si tratta di una forma di indeterminismo. Questo argomento è stato trattato in [Dragoni, 2009] perché la struttura mentale, il linguaggio e la semantica introdotte da questo articolo possono portare a una forma di indeterminismo come quella di questo tipo. Più precisamente, si è visto che l'indeterminismo può essere espresso dal formalismo che adotta quel linguaggio, ma non è stata per nulla presa in considerazione la struttura e la semantica.

Quindi ora vogliamo vedere come questa formula possa derivare dalla struttura e dalla semantica per mezzo di due esempi. Nel primo proviamo a partire dai fatti del mondo nelle attitudini mentali, e tramite le formule di reflection\_up [Dragoni et al., 2002] arrivare alla formula. Nel secondo introduciamo direttamente la formula nello stato mentale dell'agente e vediamo come viene tradotta dalle formule di reflection\_down [Dragoni et al., 2002].

Partiamo da alcune ipotesi iniziali. Rappresenteremo solamente il primo livello dello stato mentale poiché il ragionamento potrà essere esteso a tutti i livelli allo stesso modo. Introduciamo nel nostro linguaggio  $EL_i$  due preposizioni primitive **a** e **b** che rappresentano i fatti del mondo. I casi possibili generati da queste due preposizioni sono 4 e possono essere rappresentati da  $2^4=16$  funzioni di verità.

### 3. ESTENSIONE DELLA SEMANTICA

a	b	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>	F <sub>8</sub>	F <sub>9</sub>	F <sub>10</sub>	F <sub>11</sub>	F <sub>12</sub>	F <sub>13</sub>	F <sub>14</sub>	F <sub>15</sub>	F <sub>16</sub>
F	F	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
F	V	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
V	F	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
V	V	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

#### Esempio 3.1

L'agente **i** è incerto tra **a** e **b**. Quindi non sappiamo se **a** sia vera o falsa e non sappiamo se **b** sia vera o falsa. A questo punto prendiamo la funzione di verità che rappresenta l'*or* tra le due proposizioni. Questa analisi ci porta a scegliere la rappresentazione

0
1
1
1

che viene inserita nella struttura mentale:

	<b>i</b>		
<b>BE</b>	0		<b>DE</b>
	1		
	1		
	1		
<b>BI</b>			<b>DI</b>

Da questa rappresentazione della struttura mentale, possiamo derivare la forma canonica disgiuntiva in modo da ottimizzare (o minimizzare) l'informazione.

$$(\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b) \equiv a \vee b$$

A questo punto abbiamo trovato la rappresentazione in formule, quindi possiamo inserirle nella struttura multi-contestuale del nostro agente **i**.

Possiamo notare che, con le ipotesi introdotte, e con l'utilizzo delle proprietà descritte in [Dragoni, 2009]Definizione 4, non siamo riusciti ad arrivare al risultato che volevamo. L'unica cosa che abbiamo dimostrato è che questa struttura permette solamente di rappresentare l'incertezza di un agente  $i$  su un argomento  $a$ .

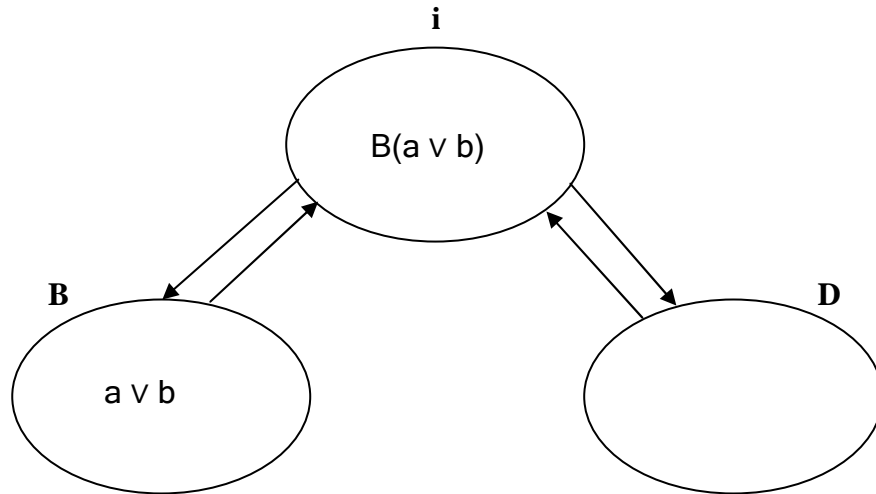


Figura 3.1

### Esempio 3.2

A dimostrazione di quanto appena detto possiamo effettuare una contro prova, questa volta, però, partiamo dallo stato mentale dell'agente  $i$ , ed inseriamo direttamente la formula. Una situazione come questa si potrebbe verificare nel caso in cui un agente viene forzato, dall'ingegnere, ad apprendere questa formula, come se si trattasse del suo inconscio.

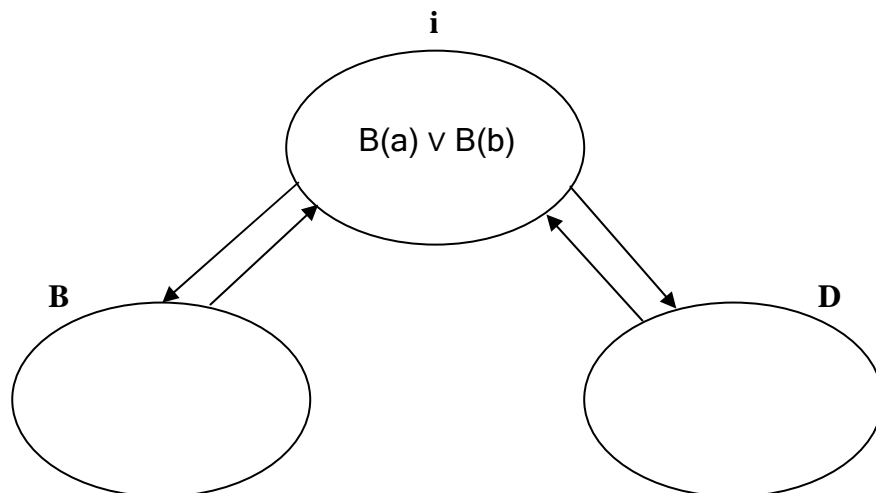


Figura 3.2

### 3. ESTENSIONE DELLA SEMANTICA

Questa rappresentazione perde ogni valore di interesse, perché non possiamo, tramite l'istruzione di *reflection down* o tramite la proprietà [Dragoni, 2009]Definizione 4, collocare i fatti all'interno dei *Beliefs* dell'agente *i*. Inoltre implica che non esiste un albero di verità capace di rappresentare un'assegnazione di questo tipo.

Abbiamo quindi dimostrato come, con la struttura descritta in [Dragoni, 2009], non si potrà mai avere una formula del tipo:

$$i: iB(a) \vee iB(\neg a)$$

anche se questa risulta essere accettabile dal linguaggio.

Ma abbiamo veramente bisogno di rappresentare questa formula nello stato mentale? La risposta è sì, ed ora lo proviamo.

#### 3.1.2 Perché abbiamo bisogno dell'indeterminismo?

Per rispondere a questa domanda introduciamo un esempio. Partiamo dalla rappresentazione vista prima in cui un agente *i* è incerto tra due proposizioni *a* e *b*.

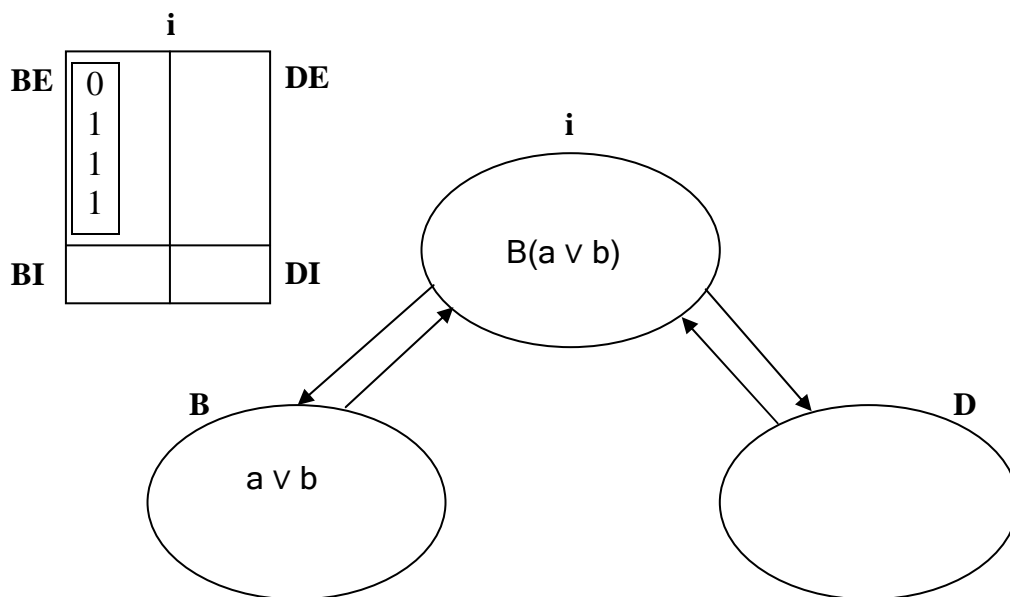


Figura 3.3

A questo punto l'agente *i*, nel caso in cui volesse risolvere l'incertezza, deve desiderare di credere soltanto in una delle due proposizione, cioè o credere in *a* o credere in *b*, ma come possiamo rappresentare questo desiderio? Se utilizziamo la struttura mentale e la semantica utilizzate in [Dragoni, 2009] possiamo solamente rappresentare il tutto introducendo un desiderio interno (Figura 3.4).

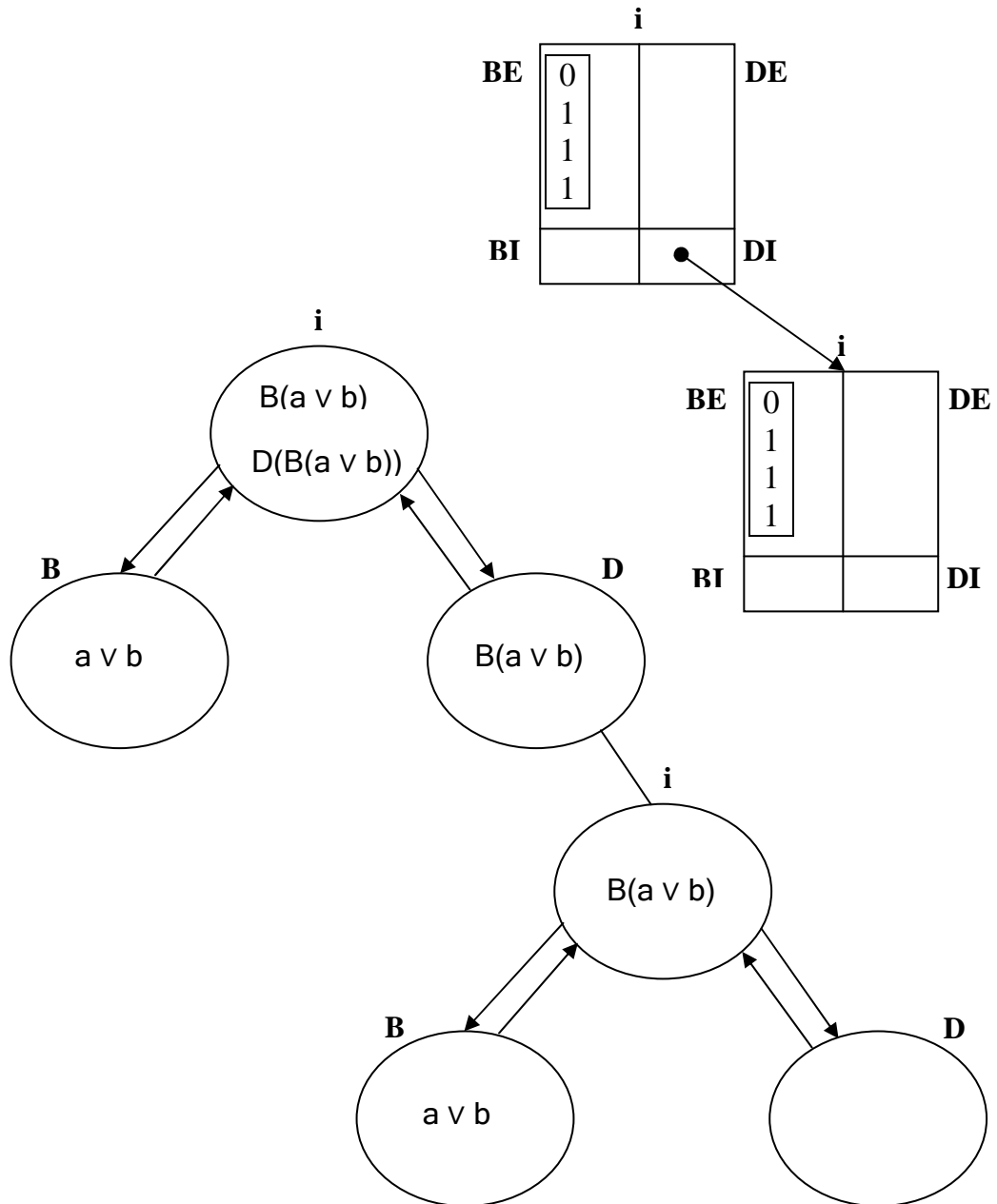


Figura 3.4

### 3. ESTENSIONE DELLA SEMANTICA

Quanto rappresentato dalla struttura mentale e dallo stato mentale dell'agente **i**, equivale a dire che l'agente desidera rimanere incerto, come esplicitato dalla formula:

$$i: iD(iB(a \vee b))$$

Quanto detto lo vediamo subito anche dalla struttura mentale, nella quale troviamo le stesse funzioni di verità, sia nel riquadro dei Beliefs sia nel riquadro dei Desires/Beliefs, cioè quello che l'agente crede è uguale a quello che l'agente desidera credere, portandolo in uno stato di equilibrio mentale.

Il risultato ottenuto non permette di arrivare alla conclusione cercata, cioè quella di rappresentare il desiderio dell'agente **i** di risolvere l'incertezza, cioè la struttura mentale e la semantica non hanno le potenzialità di rappresentare questa situazione. Ma vediamo ora in dettaglio di ciò che avremmo bisogno per risolvere il problema appena illustrato. Per far questo, quindi, abbiamo bisogno che l'agente desideri di credere in una sola interpretazione in cui **a** è vera oppure di credere in una sola interpretazione in cui **b** è vera, e queste due interpretazioni sono rappresentate dalle seguenti funzioni di verità rispettivamente

e

0
0
1
1

0
1
0
1

Nella struttura mentale che abbiamo visto fino ad ora, sono rappresentati i quattro stati **BE**, **DE**, **BI** e **DI**. I primi due riempiti con una colonna di 0 ed 1, i secondi due che si collegano alla struttura ricorsiva della rappresentazione mentale interna. La serie di 0 ed 1 rappresentano le combinazioni possibili, dei fatti del mondo, che l'agente crede vere, cioè



quelle in cui crede o quelle che lui desidera, rappresentati però da una sola colonna che indica un solo insieme di interpretazioni possibili. Ora, invece, abbiamo bisogno di far esprimere ad un agente il desiderio di credere o in una rappresentazione di interpretazioni possibili o in un'altra rappresentazione di interpretazioni possibili, perché, come visto nell'esempio, sono certo di desiderare di credere in un determinato argomento **a** o di credere in un determinato argomento **b**. La soluzione sarebbe quella di inserire all'interno delle caselle **BE** e **DE**, anche più di una colonna di 0 ed 1, più precisamente un vettore, in modo da avere più insiemi di interpretazioni..

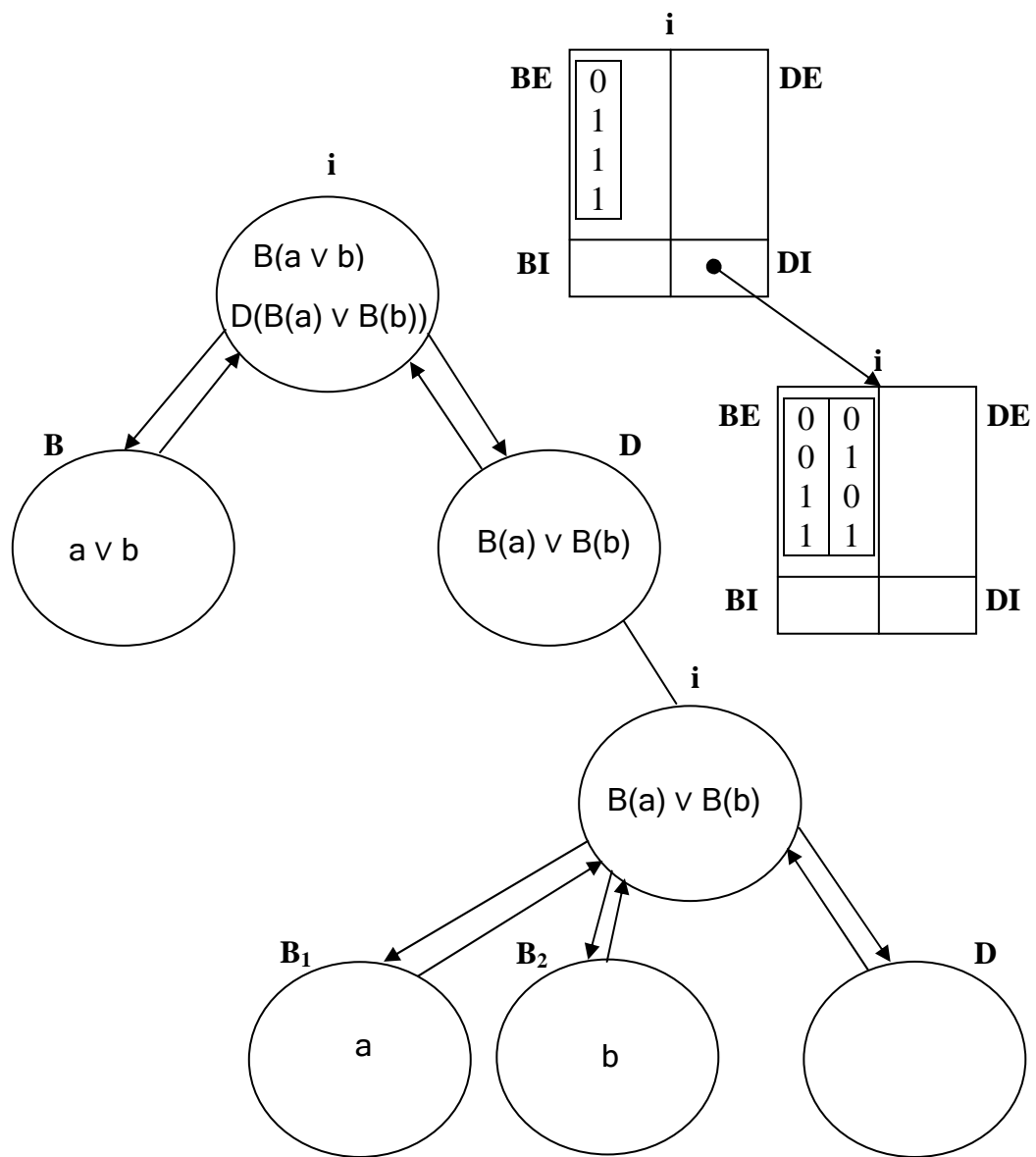


Figura 3.5

### 3. ESTENSIONE DELLA SEMANTICA

La dimostrazione può essere data dall'esempio precedente che, con l'introduzione di quanto detto, si trasforma nella struttura mentale e nella rappresentazione multi contestuale di Figura 3.5.

Come possiamo vedere, traduciamo le funzioni di verità, cioè le diverse colonne di 0 ed 1 racchiusi nella struttura mentale, in più rappresentazioni di Beliefs (o Desires), che risultano essere raggiungibili, o "potenzialmente validi" finché non si avrà la conoscenza necessaria per rimanere con un'unica rappresentazione di Beliefs (o Desires).

L'operazione di reflection up si traduce in un aggiunta dell'operatore B che racchiude i fatti per ogni Beliefs e vengono messi tutti in *or* ("V"). Naturalmente l'operatore *or* raggrupperà solamente i fatti che non sono presenti in tutti i Beliefs. Questo lo possiamo vedere con un semplice esempio

#### **Esempio 4.1**

Immaginiamo che in entrambi i contesti dello split B in Figura 3.5 avremmo avuto anche un fatto *c* creduto come vero, allora le due formule linguaggio argomentale (AL) sarebbero state:

$$*i/D/i/B_1:a \wedge c$$

$$*i/D/i/B_2:b \wedge c$$

Dalla [Dragoni, 2009]Definizione 4 più l'ipotesi attuale abbiamo che:

$$* i/D/i: (iB(a) \wedge iB(c)) \vee (iB(b) \wedge iB(c)) \equiv i/D/i: iB(c) \wedge (iB(a) \vee iB(b))$$

L'equivalenza deriva dal Teorema 2.4 presente nell'articolo [Dragoni, 2009].

In questo modo permettiamo di rappresentare il desiderio dell'agente *i* di credere una cosa oppure di credere un'altra e non il desiderio di rimanere nell'incertezza:

$$i: iD(iB(a) \vee iB(b)).$$

Con quanto introdotto possiamo anche effettuare il passaggio al contrario. Cioè, come nell'Esempio 2, inserendo una formula che contenga un "or" di due stessi operatori, e tramite l'operazione di reflection down, possiamo *splittare* i fatti in più Beliefs o in più Desires. L'esempio appena descritto non fa altro che confermare l'utilità pratica della rappresentazione vista nei riquadri 3 e 4 della Figura 2.6.

La formula appena vista, naturalmente, non è valida solo nel contesto dell'esempio sopra illustrato, ma possiamo anche prendere in esame un altro esempio. Mettiamo che un agente *i* deve rappresentare lo stato mentale di un altro agente *j*, con un suo linguaggio esterno  $EL_j$ . Nel momento in cui l'agente *i* deve rappresentare una proposizione del linguaggio  $EL_j$  che l'agente *j* conosce in termini di vero o falso, e non appartiene al linguaggio  $EL_i$ , dobbiamo sicuramente utilizzare la formula appena vista, cioè l'operatore *or* tra due formule attitudinali, in caso contrario viene rappresentata l'incertezza dell'agente *j*, che è differente dallo stato mentale dell'agente *j*.

In conclusione, arriviamo ad affermare che, la formula all'inizio di questo capitolo risulta essere vantaggiosa. Ma, in realtà, questo, non è esattamente vero. La formula vista all'inizio introduce sempre un indeterminismo, e sottolineiamo ancora l'importanza di evitare quel tipo di rappresentazione. Invece, ciò di cui abbiamo bisogno è la rappresentazione di quella formula, all'interno di un pensiero certo, sia esso Beliefs o sia esso Desires. In questo modo evitiamo l'indeterminismo poiché viene racchiuso all'interno di un determinismo, cioè all'interno di un desiderio o di una credenza di cui siamo consapevoli.

Quindi, per rendere effettivo quanto appena detto, risulta evidente che bisogna evitare di avere uno *split* al primo livello dello stato mentale o della struttura mentale, ma renderlo possibile solamente dal secondo livello in poi, senza limitazioni.

**Definizione 2.** Lo *split* di un contesto è la suddivisione in più contesti Beliefs o Desires all'interno di un unico x-module.

## 3.2 Revisione della semantica

La rappresentazione che è stata spiegata nel precedente capitolo implica un cambiamento, o meglio ancora una estensione, alla semantica descritta in [Dragoni, 2009], poiché si introduce un caso che in quel testo non era stato preso in considerazione.

**[Dragoni, 2009]Definizione 7c.** Un albero di verità, o Truth Tree,  $TT$  è l'assegnamento di uno o più insiemi di interpretazione per ogni linguaggio  $EL_j$  di ogni contesto  $*i/P$  in  $MS$  che mappi i loro corrispondenti mondi  $*i/P$  in  $\Sigma$ .

**[Dragoni, 2009]Definizione 8c.** Dato  $*i/P^{TT}$ , denota gli insiemi di interpretazioni associati dall'albero di verità  $TT$  al mondo  $*i/P$ .

*Dimostrazione.* Nel caso in cui ci sia, ad un determinato livello dell' $MS$ , un unico contesto  $P$  o uno split in più contesti  $P$ , comunque le loro rappresentazioni si trovano nel continente  $\mathbf{P}$  di  $\Sigma$ , sotto forma di un'unica colonna o di più colonne (vedere Figura 3.5).

**[Dragoni, 2009]Definizione 9c.** Un albero di verità, o Truth Tree,  $TT \in \Omega$  è un *modello* per una formula etichettata  $*i/P:\alpha$ , dove  $\alpha \in AL_i$  è una formula atomica di  $EL_i$ , se e solo se  $\alpha$  è vera in tutte le interpretazioni di almeno uno degli insiemi  $*i/P^{TT}$ , scritto  $*i/P^{TT} \models \alpha$ . Potremmo anche scrivere come  $TT \models *i/P:\alpha$ .

La definizione 9c non cambia la semantica trattata nell'articolo [Dragoni, 2009] ma ne estende le proprietà, inglobando la semantica precedente. Permette quindi di dare un significato a delle sintassi che erano possibili ma prive di validità. Con questa nuova definizione il Teorema 1b

trattato nell'articolo [Dragoni, 2009] rimane ugualmente soddisfatto perché non viene a mancare la validità dei passi nella dimostrazione del teorema.

*Dimostrazione.* Mostriamo che  $*i:iP(\alpha)\forall iP(\beta)\rightarrow *i:iP(\alpha \vee \beta)$ .

1. se  $*i:iP(\alpha)$  è vero allora  $*i/P:\alpha$  è vero ([Dragoni, 2009]Definizione 4)
2. se  $*i/P:\alpha$  è vero allora  $*i/P:\alpha\vee\beta$  è vero ([Dragoni, 2009]Definizione 9c e semantica dell'*or*) cioè è vero in tutte le interpretazioni di almeno un insieme (o contesto P)
3. se  $*i/P:\alpha\vee\beta$  è vero allora  $*i:iP(\alpha\vee\beta)$  è vero ([Dragoni, 2009]Definizione 4)
4.  $*i:iP(\alpha)\rightarrow *i:iP(\alpha\vee\beta)$  (unendo il passo 1, 2 e 3)
5.  $*i:iP(\beta)\rightarrow *i:iP(\alpha\vee\beta)$  (unendo il passo 1, 2 e 3 con  $\beta$  al posto di  $\alpha$ )
6.  $*i:iP(\alpha)\forall iP(\beta)\rightarrow *i:iP(\alpha\vee\beta)$  (dal passo 4, 5 e dalla [Dragoni, 2009]Definizione 5.2)

Inoltre proviamo che non tiene l'inverso  $*i:iP(\alpha\vee\beta)\rightarrow *i:iP(\alpha)\forall iP(\beta)$ .

1. se  $*i:iP(\alpha\vee\beta)$  è vero allora  $*i/P:\alpha\vee\beta$  è vero ([Dragoni, 2009]Definizione 4)
2. se  $*i/P:\alpha\vee\beta$  è vero allora non è vero che almeno uno dei due  $*i/P:\alpha$  o  $*i/P:\beta$  è vero, perché la [Dragoni, 2009]Definizione 9c richiede che siano vere in tutte le interpretazioni di almeno un insieme, fatto smentito dalla formula  $*i/P:\alpha\vee\beta$  che ci dice che né  $\alpha$  né  $\beta$  sono veri in tutte le interpretazioni.

La dimostrazione si ferma qui, validando solamente la prima parte e non la seconda, così come è stato dimostrato in [Dragoni, 2009].

Una cosa importante da sottolineare, è la Definizione 4 dell'articolo [Dragoni, 2009], che qui ricordiamo.

**[Dragoni, 2009]Definizione 4.** Una formula mentale atomica di  $ML_i$   $*i:iP(\delta)$ , con  $\delta \in AL_i$ , è vera se e solo se la formula attitudinale  $*i/P:\delta$  è vera

### 3. ESTENSIONE DELLA SEMANTICA

Questa definizione esplicita un legame tra formula mentale e formula attitudinale. Ma se leggiamo bene, abbiamo solamente un legame tra formula mentale atomica e formula attitudinale che può essere atomica o non atomica. Quindi, andando a prendere la Definizione 5 dell'articolo [Dragoni, 2009], cioè la definizione delle formule mentali non atomiche, non abbiamo nessun legame con la formula attitudinale. Per dimostrare questo partiamo dalla Definizione 5.1.

**[Dragoni, 2009]Definizione 5.1.** Data una formula mentale  $\alpha$  di  $ML_i$  allora  $*i:\neg\alpha$  è vera se e solo se  $*i:\alpha$  è falsa.

Per spiegare questa definizione dobbiamo prendere la formula mentale con le dovute proprietà descritte nella [Dragoni, 2009]Definizione 3, che dice che ogni formula mentale è racchiusa dagli operatori di livello (iB o iD). In questo esempio prendiamo P per non specificare in quale contesto attitudinale ci troviamo perché risulta essere equivalente. Riscrivendo la formula come abbiamo appena detto, diventerebbe che  $*i:\neg P(\alpha)$  è vera se e solo se  $*i:P(\alpha)$  è falsa. Da questa nuova rappresentazione che è equivalente a quella della [Dragoni, 2009]Definizione 5.1 possiamo dedurre che  $*i:\neg P(\alpha) \equiv *i:P(\neg\alpha)$ .

*Dimostrazione.*

1.  $*i:\neg P(\alpha)$  è vera se e solo se  $*i:P(\alpha)$  è falsa ([Dragoni, 2009]Definizione 5.1)
2. se  $*i:P(\alpha)$  è falsa allora  $*i/P:\alpha$  è falsa in tutte le interpretazioni ([1]Definizione 4)
3. se  $*i/P:\alpha$  è falsa in tutte le interpretazioni allora  $*i/P:\neg\alpha$  è vera in tutte le interpretazioni
4. se  $*i/P:\neg\alpha$  è vera in tutte le interpretazioni allora  $*i:P(\neg\alpha)$  è vera ([Dragoni, 2009]Definizione 4)
5.  $*i:\neg P(\alpha) \equiv *i/P:\neg\alpha$

Naturalmente, questa equivalenza, risulta al quanto strana. Il fatto di non conoscere la lettera proposizionale  $\alpha$ , o meglio, il fatto che l'agente  $i$  non crede/desideri  $\alpha$ , significa che l'agente  $i$  non crede/desideri che  $\alpha$  sia vera o crede/desideri che  $\alpha$  sia falsa. Questo è differente dal dire che non credere/desiderare  $\alpha$  significa dire che  $\alpha$  è falsa, come vediamo dalla dimostrazione precedente al punto 5. Quindi dobbiamo modificare la definizione 5.1 al fine di dare una giusta interpretazione al fatto di non credere/desiderare  $\alpha$ , cioè  $*i:\neg P(\alpha)$ . Modificando la definizione abbiamo che:

**[Dragoni, 2009]Definizione 5.1b.** Data una formula mentale  $*i:P(\alpha) \in ML_i$ , allora  $*i:\neg P(\alpha)$  è vera se e solo se  $*i:P(\alpha)$  è falsa e  $*i:P(\neg\alpha)$  è falsa.

*Dimostrazione.* Vediamo subito dalla dimostrazione precedente che rimangono validi tutti i punti eccetto il punto 3, partendo dalla definizione. Perché dire che  $*i:P(\alpha)$  è falsa e  $*i:P(\neg\alpha)$  è falsa equivale a dire per la [Dragoni, 2009]Definizione 4 che  $\alpha$  è falsa e che  $\neg\alpha$  è falsa. Questo significa dire che  $\alpha$  è vera non in tutte le interpretazioni di tutti gli insiemi. Siamo quindi arrivati al risultato che volevamo, cioè che  $*i:\neg P(\alpha)$  equivale a dire che  $\alpha$  non è definita all'interno del contesto attitudinale  $P$ .

Naturalmente da questa nuova definizione si può dedurre che

$$*i:\neg P(\alpha) \equiv *i:\neg P(\neg\alpha).$$

*Dimostrazione.* Poiché essendo  $*i/P: \alpha$  e  $*i/P: \neg\alpha$  non vere in tutte le interpretazioni di tutti gli insiemi, allora  $*i:P(\alpha)$  è falsa e  $*i:P(\neg\alpha)$  è falsa, quindi la loro negazione sarà vera ed equivalente.

Ora non ci resta che definire la semantica dello *split* per dare una completa e corretta interpretazione alla Definizione 5.2 dell'articolo [Dragoni, 2009].

**Definizione 3.** Dato uno *split*, presente al livello  $*i/P$  della struttura multi contestuale  $MS_{i,G}$ , rappresentato da  $n$  contesti  $P$ , dove ogni contesto

### 3. ESTENSIONE DELLA SEMANTICA

contiene un insieme di formule attitudinali  $AL_i$ . Lo *split* deve essere rappresentato nello stato mentale dall'unione in *or* di tutte le formule mentali  $ML_i$  dei contesti P. In formule avremo un'unica formula mentale:

$$*i:(P_1(..)\wedge P_1(..)\wedge.. \wedge P_1(..))\vee(P_2(..)\wedge P_2(..)\wedge.. \wedge P_2(..))\vee \\ \dots \vee(P_n(..)\wedge P_n(..)\wedge.. \wedge P_n(..)).$$

Riprendendo la definizione di *split* vista nel paragrafo precedente, capiamo che i pedici del contesto P non hanno alcun senso, poiché rimangono le proprietà di un unico contesto. Questo permette di applicare il [Dragoni, 2009]Teorema 2.4, la distribuzione di *or* su *and*, in modo da avere una rappresentazione più consona dello stato mentale. Applicando la distribuzione si avrà sempre come risultato formule mentali atomiche, per argomenti che sono veri in tutte le interpretazioni di tutti gli insiemi, e formule mentali non atomiche, cioè in *or*, per quegli argomenti che sono veri in tutte le interpretazioni ma non in tutti gli insiemi. Questo lo possiamo vedere dall'Esempio 4.1 che introduce la proprietà dello *split* appena vista.

Così, grazie alla Definizione 3 abbiamo espresso il legame tra una formula mentale non atomica e una formula attitudinale, come visto nella [Dragoni, 2009]Definizione 5.2b. Ora passiamo ad interpretare anche gli altri due punti della [Dragoni, 2009]Definizione 5.

Risulta facile ed intuitivo spiegare la [Dragoni, 2009]Definizione 5.3, poiché grazie alla semantica dell'*and* si hanno due formule mentali atomiche, quindi ci possiamo ricollegare alla [Dragoni, 2009]Definizione 4. Passiamo ora a spiegare la [Dragoni, 2009]Definizione 5.4, e aggiungendo che non si avrà mai una situazione uguale al caso illustrato.

**[Dragoni, 2009]Definizione 5.4.** Date due formule mentali  $\alpha$  e  $\beta$  in  $ML_i$ , allora  $*i:\alpha \rightarrow \beta$  se e solo se  $*i:\neg\alpha \vee \beta$  è vera.

Riscritta con le proprietà delle formule mentali [Dragoni, 2009]Definizione 3:



**[Dragoni, 2009]Definizione 5.4.** Date due formule mentali  $P(\alpha)$  e  $P(\beta)$  in MLI, allora  $*i:P(\alpha) \rightarrow P(\beta)$  se e solo se  $*i:\neg P(\alpha) \vee P(\beta)$  è vera.

*Dimostrazione.* Il fatto di avere  $*i:\neg P(\alpha) \vee P(\beta)$  è vera significa che siamo in presenza di uno split, poiché si ha un *or* tra due formule mentali. Il primo termine dell'*or* è  $\neg P(\alpha)$ . Dalla [Dragoni, 2009]Definizione 5.1b vediamo che  $\neg P(\alpha)$  è vera solo se  $\alpha$  è vera non in tutte le interpretazioni di tutti gli insiemi. Questo implica che per la Definizione 3 si ha la formula mentale  $*i:\neg P(\alpha)$  da sola.

Quindi non sarà mai possibile avere la negazione di una formula mentale in *or* con un'altra formula mentale.

Quanto appena detto risulta non essere un caso del tutto positivo, perché non è possibile creare un'implicazione tra due formule mentali, cioè tra due attitudini. Per far fronte a questo difetto viene introdotto, nel capitolo successivo, un nuovo insieme che interviene sia sulla struttura mentale sia sul sistema multi contestuale. Bisogna però differenziare un caso in cui l'implicazione tra due formule mentali può essere supportata dalla struttura. Questa situazione comprende l'implicazione tra due formule mentali dello stesso contesto attitudinale, cioè due formule mentali che hanno lo stesso percorso identificativo del contesto in cui si trovano. Possiamo dimostrare quanto detto.

**Teorema 1.** Per ogni contesto  $*i/P$ ,  $*i:iP(\alpha) \rightarrow iP(\beta) \equiv *i:iP(\alpha \rightarrow \beta)$

*Dimostrazione.* La dimostrazione vuole far vedere che entrambe le formule del teorema portano allo stesso risultato. Partiamo dal primo caso della dimostrazione, cioè l'ipotesi che  $*i:iP(\alpha) \rightarrow iP(\beta)$ :

1. se  $*i/P:\alpha$  è vero allora  $*i:iP(\alpha)$  è vero ([Dragoni, 2009]Definizione 4)
2. se  $*i:iP(\alpha)$  è vero allora  $*i:iP(\beta)$  è vero (dall'ipotesi iniziale)

### 3. ESTENSIONE DELLA SEMANTICA

3. se  $*i:iP(\beta)$  è vero allora  $*i/P:\beta$  è vero ([Dragoni, 2009]Definizione 4)
4. se  $*i:\alpha$  è vero allora  $*i:\beta$  è vero (unendo il passo 1 con il 3)  
in conclusione: se  $*i:iP(\alpha)$  è vero allora anche  $*i:iP(\beta)$  è vero.

Dimostriamo ora il secondo caso, cioè  $*i:iP(\alpha \rightarrow \beta)$ :

1. dato  $*i:iP(\alpha \rightarrow \beta)$  allora  $*i/P:\alpha \rightarrow \beta$  ([Dragoni, 2009]Definizione 4)
2. se  $*i/P:\alpha$  è vero allora  $*i:iP(\alpha)$  è vero ([Dragoni, 2009]Definizione 4)
3. inoltre se  $*i/P:\alpha$  è vero allora  $*i/P:\beta$  è vero (dall'ipotesi iniziale e dal passo 1)
4. se  $*i/P:\beta$  allora  $*i:iP(\beta)$  è vero ([Dragoni, 2009]Definizione 4)  
in conclusione: se  $*i:iP(\alpha)$  è vero allora anche  $*i:iP(\beta)$  è vero.

Una cosa da sottolineare e da non trascurare è la parte del teorema che indica ogni contesto  $*i/P$ . Quindi si vuole focalizzare l'attenzione, non solo sul fatto che il teorema è valido per ogni percorso, ma soprattutto si vuole vincolare il teorema riferendosi ad un unico contesto  $P$ , in modo da rispettarne la validità.

## 4 INSIEME DEI VINCOLI DI RAGIONAMENTO

---

## 4. INSIEME DEI VINCOLI DI RAGIONAMENTO

### 4.1 Natura dei vincoli

Come è stato detto nel precedente capitolo, l'implicazione logica tra due linguaggi mentali, cioè tra due attitudini, non ha possibilità di essere rappresentata nel sistema multi contestuale, così come nella struttura. Ma volendo sfruttare questa espressione nel migliore dei modi, viene creato un insieme che permette di obbligare delle credenze e dei desideri a partire dalla verità di altre credenze o desideri. Questo è l'insieme dei **vincoli di ragionamento** ed impone delle restrizioni sullo stato mentale dell'agente. Queste restrizioni agiscono piuttosto sul modo di ragionare dell'agente. Cioè, le credenze o i desideri di un determinato fatto implicano un'ulteriore credenza o un'ulteriore desiderio. Quindi, quest'ultimo enunciato deve avere, naturalmente, un connesso logico, che viene rappresentato dal vincolo di ragionamento.

Un classico esempio può essere quello del pescatore che vedendo il mare mosso, decide di non salpare con la barca. In formule e attitudini mentali avremo:

$$B(\text{mosso}(\text{mare})) \rightarrow D(\neg\text{partenza}(\text{barca}))$$

Che cosa vuol dire una rappresentazione di questo tipo? Il fatto del mondo che rappresenta il mare mosso, nel momento che risulta essere vero, è rappresentato nel contesto Beliefs dello stato mentale del pescatore. Questo implica l'interpretazione vera del fatto 'non salpare con la barca', nel contesto Desires dello stato mentale del pescatore.

Ora prendiamo una situazione in cui nello stato mentale dell'agente troviamo  $D(\neg\text{partenza}(\text{barca}))$ , questo di per sé è possibile, senza esserci nessuna spiegazione dietro. Con questo vogliamo dire che in una comunicazione tra più agenti, non c'è nessuna spiegazione sul perché l'agente non desidera salpare con la barca. Invece, con l'introduzione del vincolo di ragionamento visto sopra, il desiderio può essere spiegato. Intanto sottolineo il fatto che non verranno utilizzati termini come 'tesi', 'teorema' e 'ipotesi' proprio per evitare di dare riferimenti all'implicazione

logica che si vuole evitare in questo contesto, e potremmo scrivere un intero saggio sul perché di questa scelta, partendo da Aristotele fino a Bertrand Russell. Lascio solo un indizio, per chi volesse cimentarsi in questa scelta, che è quello di utilizzare un'ulteriore rappresentazione della formula vista sopra e legare la [Dragoni, 2009]Definizione 5.1b con la semantica dell'*or*.

Tornando al discorso principale, il vincolo viene preso in considerazione da sinistra verso destra nel momento in cui è vero il termine di sinistra che implica il vero per il termine di destra. Inoltre, si prenderà in considerazione il vincolo da destra verso sinistra solamente se sono veri entrambi i membri, sia di destra che di sinistra. Questo è facilmente spiegabile prendendo l'esempio precedente del pescatore. Il desiderio di non salpare può essere vero e scaturire da infiniti fattori, anche mentre il fatto che 'il mare è mosso' è falso. Però nel momento in cui il desiderio di non salpare è vero ed anche il fatto che il mare è mosso è vero, sussiste il vincolo di ragionamento che c'era il legame tra questi due fatti dello stato mentale dell'agente. Quindi il desiderio di non salpare è scaturito solamente dal mare mosso.

## 4.2 Semantica

L'insieme dei vincoli di ragionamento non è vincolato a rimanere uguale per tutta la vita dell'agente, ma sarà anche esso partecipe dell'evoluzione dello stato mentale dell'agente in corrispondenza dell'interazione con l'ambiente esterno, o meglio con l'interazione tra altri agenti o persone. L'insieme non rappresenterà solamente i vincoli dello stato mentale dell'agente, ma avrà anche una rappresentazione dei vincoli di ragionamento di tutti gli agenti del gruppo. Ciò è possibile perché i vincoli di ragionamento sono legati allo stato mentale dell'agente, quindi avendo una rappresentazione di tutti gli stati mentali degli agenti, si può ricreare anche il loro insieme di vincoli. Come più volte detto, l'insieme dei vincoli di ragionamento è legato allo stato mentale dell'agente, e più appropriatamente rimane legato al primo livello dello stato mentale

#### 4. INSIEME DEI VINCOLI DI RAGIONAMENTO

dell'agente. E' facile intuire che i vincoli di ragionamento che riguardano diversi livelli dello stato mentale, o che riguardano altri agenti del gruppo, sono rappresentati dal percorso dell'albero che è univoco. Questa ultima precisazione risulta essere molto vantaggiosa all'interno della struttura innovativa che stiamo creando perché permette di avere all'interno dell'insieme dei vincoli di ragionamento situazioni che presentano vincoli tra differenti livelli di attitudini mentali o addirittura tra differenti agenti.

L'insieme dei vincoli di ragionamento risulterà legato sia alla struttura mentale sia al sistema multi-contestuale. Nel sistema multi-contestuale troveremo l'insieme dei vincoli rappresentato con il simbolo  $\Delta$ . Questo insieme sarà composto solamente da formule mentali anche se i vincoli di ragionamento agiscono sul linguaggio attitudinale. Il tutto può essere spiegato con due definizioni.

**Definizione 4.** L'insieme dei vincoli di ragionamento  $\Delta_{i,G}$ , dove  $i$  è l'agente e  $G$  è il gruppo a cui appartiene, racchiude coppie di formule mentali atomiche legate dal simbolo  $\rightarrow$ .

**Definizione 5.** Date due formule  $*i_1:P(\alpha)$ ,  $*i_2:P(\beta) \in ML_i$ , dove 1 e 2 indicano due percorsi diversi, legate dal simbolo  $\rightarrow$ , allora la formula  $*i_1:P(\alpha) \rightarrow *i_2:P(\beta) \in \Delta_{i,G}$  implica che nel momento in cui in  $*i_1/P^{TT}_1$  esiste una funzione di verità rappresentata da  $TT_1$  che rende vero  $\alpha$  allora deve esserci in  $*i_2/P^{TT}_2$  una funzione di verità rappresentata da  $TT_2$  che rende vero  $\beta$ .

Come detto in precedenza, dalla Definizione 5 vediamo che l'insieme dei vincoli di ragionamento agisce sia sulla struttura mentale sia sul sistema multi contestuale.

Con l'introduzione dell'insieme dei vincoli introduciamo dunque una dinamica dello stato mentale dell'agente, perché l'albero delle funzioni di

verità subisce una trasformazione che fa evolvere lo stato della conoscenza dell'agente.

### **4.3 Conclusioni**

Quindi l'introduzione dei vincoli di ragionamento ci spinge sempre più verso l'obiettivo principale di questa struttura. Quest'ultima ha come scopo quello di riuscire a rappresentare lo stato mentale di diversi agenti che interagiscono fra di loro. Inoltre deve fare in modo di rappresentare in maniera appropriata e il più chiara possibile i legami tra i diversi ragionamenti che effettuano i singoli agenti. Il fatto di avere una chiara e semplice rappresentazione di un gruppo di agenti può portare a vantaggi grandiosi nel momento in cui leghiamo questa struttura ad un caso reale, ad esempio si può pensare ad un ambiente internazionale in cui gli agenti rappresentano diversi capi di stato, per riuscire a capire come agiscono e a che scopo agiscono, nel momento in cui interagiscono tra di loro. Un altro esempio può essere quello dei concorrenti nel campo industriale, e così via, pensando a infiniti casi di studio che richiedono l'interazione tra diverse persone che non sono portate a motivare in maniera chiara e completa le proprie azioni.

#### 4. INSIEME DEI VINCOLI DI RAGIONAMENTO



## 5 COMUNICAZIONE TRA AGENTI

---

### 5.1 Introduzione

Quindi grazie a questa struttura dobbiamo vedere come evolvono le dinamiche di interazione tra un gruppo di agenti, per modellare i loro stati mentali. Come più volte detto queste dinamiche partono dall'interazione tra agenti che avvengono per mezzo di **speech act** (o atti comunicativi) che permettono la comunicazione tra agenti.

Per riuscire ad arrivare ad una perfetta rappresentazione degli speech act è nato proprio un settore di studio dei linguaggi di comunicazione tra agenti, o anche detto Agent Communication Language (ACL). In questa tesi non è stato sviluppato nessun protocollo per la comunicazione tra agenti, ma riportiamo comunque la teoria che sta alla base dell'atto comunicativo, perché è questo che permette l'evoluzione dello stato mentale. In un ambiente multi-agente, l'alto grado di interazione richiede che l'agente sia capace di scambiare informazioni usando un linguaggio di alto livello e protocolli di comunicazione. Riferendoci alla classica teoria degli speech act, sembra che l'idea di predizione degli effetti di un'espressione sullo stato mentale del destinatario sia piuttosto ottimista. Questi effetti non sono completamente predicibili a priori da parte di chi parla. Con alcuni effetti principali prevedibili ci sono altri effetti secondari imprevedibili. Questo può essere visto come il risultato di un qualche tipo di plausibile inferenza, di particolare abduzione, effettuata dall'ascoltatore sulla comunicazione ricevuta, rispetto al suo stato mentale e rispetto alla precedente immagine mentale che lui ha dello stato mentale del parlante.

### 5.2 Agent Communication Language

Negli anni recenti la ricerca in ACL è stata uno dei principali sotto settori della ricerca in agenti autonomi e sistemi multi-agente. Una delle idee basi principali della ricerca in ACL è un linguaggio di alto livello e di

tipo astratto. Nel campo dei calcolatori distribuiti, la comunicazione è generalmente vista come un'entità di scambio di messaggi, che coordinano il loro comportamento attraverso l'uso di protocolli di comunicazione [Sharp, 2000] [Bochmann and Sunshine, 1980]. Nel campo dell'intelligenza artificiale distribuita (oggi più riferita ai sistemi multi-agente) gli agenti di comunicazione sono visti come entità autonome, ad alto livello e eterogenee che ingaggiano dialoghi, conversazioni e negoziazioni con ogni altro agente, per coordinare il loro comportamento [Weiss, 2000] [Jennings et al., 2001] [Flores and Kremer, 2001] [Labrou and Finin, 1998]. La ricerca in agenti e sistemi multi-agente è caratterizzata dall'essere inerentemente multidisciplinare ed unisce le idee di molti campi scientifici tradizionali:

- Intelligenza artificiale – ingegneria della conoscenza, ragionamento epistemico, pianificazione, pro-attitudini, orientamento agli obiettivi, intelligenza, autonomia, ontologie, logica modale ecc.
- Sistemi distribuiti e real-time – progettazione dei protocolli, concorrenza, parallelismo, timing ecc.
- Ingegneria del software – metodologie formali, metodologie orientate agli oggetti;
- Linguistica, filosofia e scienze sociali – semiotica, teoria degli speech act, analisi delle conversazioni, epistemologia, negoziazioni e teoria dei lavori di gruppo ecc.

Gli agenti sono progettati per collaborare autonomamente con altri agenti in modo da soddisfare, entrambi, i loro obiettivi interni; la condivisione esterna delle richieste è controllata da una società multi agente – così come per gli umani. Il loro lavoro deve essere o collaborativo (lavoro di gruppo) o di natura competitiva.

Il campo di ricerca in ACL può essere suddiviso in tre differenti aree (che sono comunque connesse tra di loro):

1. Approccio mentale e sociale
2. Relazione tra atti comunicativi (speech act) e conversazione
3. Domini di applicazione

## 5. COMUNICAZIONE TRA AGENTI

L'approccio tradizionale che definisce la semantica della comunicazione tra agenti è basato su una prospettiva mentalistica dell'intelligenza artificiale, come ad esempio KQML [Labrou, 1996], FIPA ACL [FIPA], e Cohen/Levesque [Cohen and Levesque, 1990a]. Questi approcci sono concentrati sull'aspetto cognitivo della comunicazione, cioè le credenze, i desideri e le intenzioni di chi parla e di chi ascolta. Tutti questi utilizzano l'idea della teoria degli speech act [Bach and Harnish, 1979] [Searle, 1969] nella definizione della loro semantica ACL. Questi approcci tipicamente utilizzano una specie di logica modale per specificare la semantica degli speech act, un esempio è la logica epistemica [Halpern et al., 1995] [Genesereth and Nilson, 1986]. Uno dei principali motivi di questo approccio mentalistico è di capire la connessione tra le architetture degli agenti e i linguaggi di comunicazione tra agenti. Come detto fino ad ora l'architettura di agenti più popolare è quella BDI [Rao, 1996] [Georgeff et al., 1999] [Jennings and Wooldridge, 1996]. Per questo molti ricercatori si sono focalizzati sulla combinazione della semantica di ACL con le architetture di agenti BDI [Cohen and Levesque, 1995] [Cohen and Levesque, 1990b].

L'approccio sociale vede la comunicazione come un fenomeno pubblico sociale che non deve essere ridotto alla nozione mentalistica come credenze e intenzioni [Singh, 1998]. Invece suggerisce di usare concetti sociali, come obblighi, impegni, norme, convenzioni ecc. [Singh, 2000a] [Dignum and Weigand, 1995]. Questo significa che questi concetti si riferiscono ai rapporti esterni fra gli agenti (e non interni nella mente dell'agente). Obblighi e norme sono tipicamente controllate da un gruppo di agenti autonomi per coordinare le loro interazioni e comportamenti in un dato contesto sociale. L'approccio sociale è motivato da un numero di fattori, ad esempio l'uso di agenti in ambienti aperti e eterogenei come l'e-business e la logistica [Singh, 2000b]. Ci riferiamo all'articolo [Pedersen, 2001] per una discussione dei concetti di agenzia mentalistica e sociale.

Un altro problema centrale nella corrente ricerca in ACL è di definire la relazione tra speech act e conversazioni [Greaves et al., 1999].

Diversi approcci riconoscono che la teoria degli speech act forma una buona base per la definizione di ACL per le seguenti ragioni:

- La teoria degli speech act riguarda la pragmatica del linguaggio umano, cioè come realmente il linguaggio è utilizzato dagli esseri umani nella vita quotidiana. Poichè gli umani sono inoltre agenti autonomi, e dovrebbero poter comunicare con agenti artificiali, è ragionevole assumere che gli agenti artificiali comunicheranno utilizzando gli stessi principi base degli esseri umani.
- Gli agenti autonomi tipicamente comunicano per realizzare azioni, ad esempio emettendo una richiesta, lo speaker (o parlante) compie l'azione intenzionalmente per ottenere un messaggio di risposta dall'agente ricevente.

Tipicamente la semantica degli speech act basata su ACL definisce i diversi messaggi ACL come strutture atomiche, ad esempio KQML e FIPA ACL. Dall'altro lato, le conversazioni sono definite in termini di protocolli, ad esempio macchine a stati finiti, Petri Nets [Cost t al., 2000], CSP [Pedersen, 2001] ecc. Questo ha portato ad un mancato collegamento tra la semantica degli speech act e la semantica delle conversazioni. E' stato discusso che questo approccio attuale scarta l'ACL e che gli speech act, in questo caso, possono essere sostituiti con un arbitrario insieme di token come tipi di messaggi [Singh, 2000a]. Per questo è stato suggerito che la semantica degli speech act dovrebbe anche specificare come le conversazioni dovrebbero emergere da individuali speech act, ad esempio facendo qualche tipo di semantica compositiva. In questo modo l'agente potrebbe comporre tutti i tipi di conversazioni strutturate diversamente usando lo stesso linguaggio primitivo, senza aver bisogno di accordarsi sul tipo di protocollo per ogni tipo di conversazione [Flores and Kremer, 2001]. Questo tipo di semantica può fornire grande flessibilità e autonomia, rispetto all'approccio tradizionale.

Inoltre è stato suggerito l'utilizzo di politiche di conversazione invece dei protocolli. Le politiche di conversazione sono definite come

## 5. COMUNICAZIONE TRA AGENTI

principi pubblici che costringono la natura e lo scambio di speech act semanticamente coerenti fra agenti. In [Greaves et al., 1999] si definiscono le politiche di conversazione “a grana fine”, cioè i diversi vincoli sono presunti per richiamare soltanto una singola caratteristica di una particolare conversazione. Ciò è differente dai meccanismi correnti di descrizione del protocollo di conversazione, il quale cerca di definire ogni proprietà rilevante di una conversazione con un singolo protocollo, ad esempio come le macchine a stati finiti o Petri Nets. Greaves afferma che questi meccanismi dovrebbero solo essere utilizzati per implementare particolari politiche di conversazione e non per specificarle. Le politiche di conversazione possono avere differenti gradi di intensità per limitare l’uso di ACL, in base al tipo di interazione a cui l’agente partecipa

Si è anche affermato che gli impegni sociali (o obblighi) dovrebbero essere la nozione base nella semantica degli speech act e delle conversazioni, cioè le conversazioni sono costruite su azioni basate sugli obblighi dell’agente per azioni future, ad esempio l’azione di replica ad una domanda [Singh, 2000a] [Colombetti, 2000]. Gli obblighi sono definiti come l’impegno per una linea di condotta presa da un agente relativamente ad un altro agente.

L’ultima area è il dominio di applicazione. Questo settore problematico si occupa della funzione specifica del dominio di applicazione dell’ACL. In generale i ricercatori hanno lo scopo di progettare un linguaggio che possa, almeno in teoria, essere applicabile in un vasta gamma di domini di applicazione, ad esempio nell’e-business, logistica, robotica, progettazione di HMI (Human Machine Interface) ecc.

### 5.3 Teoria degli speech act

Tradizionalmente la teoria linguistica è suddivisa in tre gruppi: sintassi, semantica e pragmatica; riferendosi a volte alla semiotica [Winograd, 1987].

La sintassi è interessata alla struttura delle forme visibili del linguaggio. Le regole sintattiche determinano il modo in cui gli elementi linguistici sono messi insieme per formare frasi o clausole.

La semantica si occupa del significato delle lingue, ad esempio le parole e il significato delle strutture di linguaggio composite. La semantica è solitamente data dal legame del costrutto sintattico in qualche dominio semantico, come ad esempio il significato della logica proposizionale è dato dai valori vero e falso.

La pragmatica riguarda il risultato dell'uso del linguaggio. La pragmatica è forse il più difficile aspetto linguistico da definire precisamente. Gli aspetti pragmatici sono spesso difficili da formalizzare, ad esempio tutti i fenomeni di lingua che non possono essere classificati come sintassi o semantica sono pragmatici. La ragione perché la pragmatica è dura da capire e da formalizzare è che tratta il linguaggio e il suo utilizzatore in un contesto sociale:

*“Pragmatics studies the use of language in human communication as determined by the conditions of society.”*[Mey, 2001].

*“La pragmatica studia l'uso della lingua nella comunicazione umana determinata dagli stati della società.”*

La pragmatica deve considerare gli aspetti cognitivi, sociali e culturali della comunicazione. E' stato discusso se la pragmatica veramente

## 5. COMUNICAZIONE TRA AGENTI

rappresenti una suddivisione in aggiunta alla sintassi e alla semantica, o se non dovrebbe essere considerata come una prospettiva differente.

La teoria degli speech act è interessata soprattutto al risultato pragmatico del linguaggio, ad esempio l'uso del linguaggio della persone piuttosto che la forma. Questa teoria originata da John L. Austin e dalla sua collezione di note lettrali *How To Do Things With Words* (1962) [Austin, 1962], nel quale suggerisce una nuova prospettiva sul linguaggio: la prospettiva del linguaggio/azione. Questa teoria è stata poi definita in modo più formale, e nominata speech act, dal filosofo John Searle nel suo libro *Speech Act* (1969) [Searle, 1969] e *Foundations of Illocutionary Logic* (1985) [Searle and Vanderveken, 1985]. Altri linguisti e filosofi hanno provato anche a combinare la teoria degli speech act con altri rami della pragmatica, così come la struttura Gricean.

L'idea base dietro la teoria degli speech act è quella di considerare il linguaggio come uno strumento per trasmettere azioni. Si sente spesso dire la seguente frase:

*“Devi fare quello che tu dici.”* (1)

In questa frase è implicato che le parole dette non effettuano nessun tipo di azione. La teoria degli speech act suggerisce che nella maggior parte delle situazioni le nostre parole realizzino realmente azioni appena le emettiamo. Prendiamo in considerazione una situazione in cui (1) potrebbe essere usata. Se un agente *a* per esempio dice ad un altro agente *b*:

*“Arrivo e ti do una mano a spostare i mobili”* (2)

ma se l'agente *a* non viene e non aiuta allora *b* potrebbe legittimamente dire (1) ad *a* o più precisamente potrebbe dire:

*“Devi fare **fisicamente** quello che tu dici”* (3)



Significa che non è sufficiente promettere di aiutare qualcuno, cioè si deve mettere in pratica fisicamente quello che si dice. Quello che la teoria degli speech act suggerisce è che questa situazione è creata dall'esecuzione dell'atto comunicativo di *a*. L'atto comunicativo di fare una promessa crea quindi un obbligo sociale del parlante *a*, verso l'ascoltatore *b*. L'atto comunicativo può anche portare ad altre azioni: *a* esprime l'intenzione di aiutare *b*, il quale potrebbe creare una credenza in *b* che *a* veramente intenda aiutarlo.

Prendiamo ora in considerazione alcuni esempi intuitivi dove una frase può portare a delle azioni. Se un agente *a* per esempio dice:

*“Io battezzo questa nave con il nome Moby.”* (4)

Prima che *a* facesse questa dichiarazione, la nave era senza nome, ma dopo la dichiarazione la nave ha un nome. Nessuna azione fisica è stata eseguita da *a*. La frase in se stessa contiene un'azione che è eseguita quando la frase viene espressa sotto giuste circostanze. Un capo dice alla sua segretaria:

*“Per favore dammi le statistiche del secondo trimestre.”* (5)

Facendo questa richiesta il capo incita la segretaria a realizzare alcune azioni in suo favore. Questo tentativo può essere visto come un'azione, un'azione del linguaggio. In fatti la maggior parte dei politici, dei giudici e degli insegnanti sono interessati a realizzare azioni del linguaggio. In queste professioni il linguaggio è il principale strumento per esprimere idee, trasportare azioni, asserire fatti ecc. Un altro esempio dal mondo finanziario:

*““Irrational exuberance and unduly escalating stock prices.” These seven simple words describing the stock market in a speech by the Chairman of the Federal Reserve, Alan Greenspan, sent markets around the world into a sharp downward spiral.”* (7)

## 5. COMUNICAZIONE TRA AGENTI

*“Esuberanza irrazionale e prezzi azionari eccessivamente crescenti.” Queste sette semplici parole che descrivono il mercato azionario in un discorso del presidente della Federal Reserve, Alan Greenspan, hanno portato i mercati di tutto il mondo in una brusca spirale di discesa:”*

Questo esempio permette di comprendere la forza delle parole. Una singola frase espressa dal “giusto” agente, al posto “giusto”, nel momento “giusto” può portare a drammatiche azioni. Un altro aspetto importante da notare degli speech act è che possono essere eseguiti su larga scala nella società, da comunicazioni globali fatte da politici, uomini d'affari ecc., a piccole conversazioni tra bambini.

### **5.4 Conclusioni**

Questo capitolo introduce solamente l'argomento della comunicazione tra agenti, e non entra in dettaglio nell'argomento. In quanto non è obiettivo di questo studio, ma serve a comprendere le dinamiche della comunicazione tra agenti, al fine di modellare la rappresentazione dello stato mentale degli agenti stessi. Si vuole solamente accennare alla ricerca che c'è dietro questo aspetto e alla base da cui è partita questa ricerca, la teoria degli speech act. Questa teoria viene solamente descritta dal punto di vista letterale, spiegando in modo semplice cosa significa una azione del linguaggio. Un introduzione molto dettagliata degli speech act può essere trovata in [Pedersen, 2002].

## 6 DINAMICA DELLA STRUTTURA MENTALE

---

### 6.1 Introduzione

Tutto quello che è stato scritto in questa tesi fino ad ora, in aggiunta al lavoro svolto da Dragoni, crea delle basi stabili per questo studio innovativo. Quindi questa struttura permette di rappresentare gli stati mentali di un agente intelligente in un gruppo di agenti essi stessi intelligenti. Arrivati a questo punto abbiamo una sintassi e una semantica che sostengono le nostre tesi. Passiamo ora a lavorare al nostro scopo principale, cioè alla problematica da cui siamo partiti: la modellazione degli stati mentali.

Come più volte ribadito in questa tesi, la struttura offre grandi potenzialità nell'analizzare, o meglio nel rappresentare, gli stati mentali di diversi agenti in un gruppo. Il gruppo non è altro che una rappresentazione di un ambiente reale in cui ci sono diversi attori, legati tra di loro per qualche fine, ad esempio in ambiente politico per vedere le diverse scelte dei capi di stato, in ambito militare per vedere le diverse strategie, in ambito industriale tra diversi concorrenti, e così via. All'interno del gruppo gli agenti interagiscono tra di loro per mezzo di speech act, che hanno lo scopo di effettuare azioni di linguaggio, come visto nel capitolo precedente. In altre parole vogliamo vedere effettivamente il perché gli agenti interagiscono tra di loro, quindi approfondire il legame che c'è tra lo stato mentale e l'atto comunicativo.

Come possiamo vedere questo legame? Se si pensa alla nostra vita quotidiana, spesso effettuiamo dialoghi con altre persone. Questi dialoghi sono spesso eseguiti al fine di arrivare ad uno scopo che abbiamo solo nella nostra testa, ma che non sempre vogliamo rendere esplicito al nostro destinatario. Quindi significa che l'interazione con un'altra persona dipende dalle idee che ho nella mia testa, o sarebbe meglio dire, che dipende dal mio stato mentale. E' ovvio pensare che se interagisco con una determinata persona, al fine di raggiungere il mio scopo, è perché conosco parzialmente ciò che ha nella testa la persona destinataria, o sarebbe meglio dire, che ho una rappresentazione dello stato mentale del destinatario. Se leghiamo quanto detto in questo momento con la nostra struttura mentale, capiamo

subito che possiamo rappresentare tutto questo scenario, in modo semplice e chiaro. Quindi grazie alla struttura mentale possiamo modellare uno scenario che vede diversi agenti interagire tra di loro. Ritornando alla domanda posta poco sopra, capiamo che il legame tra lo stato mentale e l'atto comunicativo è legato sia allo stato mentale dell'agente mittente (o speaker) sia alla rappresentazione che l'agente parlante ha dello stato mentale dell'agente destinatario (o ascoltatore).

Fatta questa introduzione, possiamo tornare al campo della nostra ricerca. In definitiva si vuole arrivare a capire le precondizioni e gli effetti di un atto comunicativo. Quindi si vuole partire dal capire quali sono le precondizioni che portano allo speech act, cioè tramite questa struttura riuscire a capire quando un agente effettuerà una comunicazione verso un altro in base alla rappresentazione del suo stato mentale. Sempre in base alla rappresentazione dello stato mentale, si vuole capire il contenuto dello speech act. Infine si vuole vedere come viene modellato lo stato mentale del destinatario nel momento in cui riceve lo speech act.

L'esecuzione degli speech act tra diversi agenti porta quindi ad una evoluzione degli stati mentali degli agenti stessi. La ricezione di uno speech act ha come conseguenza sia la modellazione dello stato mentale dell'ascoltatore, sia la modellazione dello stato mentale che l'ascoltatore ha dello speaker. Inoltre lo speaker, nel momento in cui effettua lo speech act, modella il suo stato mentale relativo all'ascoltatore in modo da predire l'impatto che lo speech act avrà, in base al suo modello.

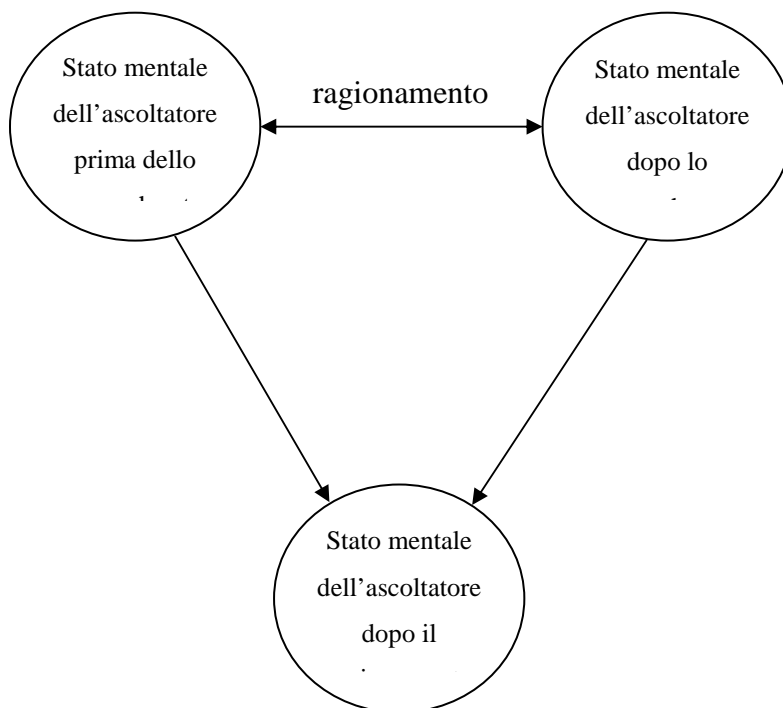
## **6.2 Linee guida**

In questa tesi non abbiamo creato un implementazione di questa struttura, compito che sarà sicuramente assegnato successivamente per un'altra tesi di ricerca. Naturalmente con un implementazione si può vedere meglio come lo stato mentale di un agente evolve. Qui diamo solo delle

## 6. DINAMICA DELLA STRUTTURA MENTALE

linee guida verso una implementazione della struttura per vedere le dinamiche e l'impatto degli speech act.

Per vedere le dinamiche degli stati mentali non bisogna solamente soffermarsi sul contenuto dello speech act, ma piuttosto bisogna capire cosa lo ha generato. Quindi risulta importante, da parte di un ascoltatore, ragionare non solamente sulla nuova rappresentazione dello stato mentale, che è stato modellato dalla ricezione dello speech act, ma bisogna effettuare un ragionamento triangolare. Cioè risulta importante prendere in considerazione nel ragionamento sia lo stato mentale che si aveva prima di ricevere lo speech act, sia lo stato mentale che si ha dopo aver ricevuto lo speech act, ed infine da questi due crearsi il nuovo modello dello stato mentale (Figura 6.1).



**Figura 6.1**

Il fatto di non effettuare il ragionamento solamente sul nuovo stato mentale modellato dallo speech act, risulta particolarmente intuitivo, poiché l'ascoltatore deve capire qual è l'intento dello speaker con l'atto comunicativo. Quindi lo speaker avrà effettuato lo speech act in base al suo

modello dello stato mentale che aveva prima, quindi anche da parte dell'ascoltatore ci deve essere l'accortezza di accorgersi delle differenze tra i due modelli di stati mentali. L'assunzione di effettuare questo tipo di ragionamento deriva dal fatto che la rappresentazione che un agente ha degli stati mentali degli altri agenti non è sempre fedele alla realtà. Quindi molte volte tramite gli atti comunicativi si riesce ad avere una rappresentazione più affidabile degli stati mentali.

## 6. DINAMICA DELLA STRUTTURA MENTALE



# BIBLIOGRAFIA

---

## BIBLIOGRAFIA

[Austin, 1962] Austin J. L., *How to do Things with Words*. Clarendon, Oxford, UK., 1962.

[Bach and Harnish, 1979] Bach K., and Harnish R. M., *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press, 1979.

[Ball et al., 1997] Ball G., Ling D., Kurlander D., Miller J., Pugh D., Skelly T., Stanskosky A., Thiel D., Van Dantzich M., and Wax T., Lifelike Computer Characters: The Persona Project Microsoft. In Jeffrey M. Bradshaw, editor, *Software Agents*, AAAI Press / The MIT Press, 1997, pp. 191-222.

[Bochmann and Sunshine, 1980] Bochmann G. V., and Sunshine C. A., Formal Methods in Communication Protocol Design. *IEEE Transactions on Communications*, COM-28, 1980, pp. 624-631.

[Bratman, 1987] Bratman M. E., *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[Bratman et al., 1988] Bratman M. E., Israel D., and Pollack M. E., Plans and resource-bounded practical reasoning. *Computational intelligence*, Vol. 4, 1988, pp. 349-355.

[Chess et al., 1994] Chess D., Harrison C., and Kershbaum A., Mobile Agents: Are They a Good Idea? Technical Report RC 19887, IBM Research Division, Yorktown Heights, new York, 1994.

[Cohen and Levesque, 1990a] Cohen P. R., and Levesque H. J., Performatives in a Rationally Based Speech Act Theory. In *Meeting of the Association for Computational Linguistics*, 1990, pp. 79-88.

[Cohen and Levesque, 1990b] Cohen P. R., and Levesque H. J., Intention is Choice with Commitment. *Artificial Intelligence*, Vol. 42, N. 3, 1990, pp. 213-261.

[Cohen and Levesque, 1995] Cohen P. R., and Levesque H. J., Communicative Actions for Artificial Agents. In *Proc. ICMAS-95*, June 1995, AAAI Press, pp. 65-72.

[Colombetti, 2000] Colombetti M.. Commitment-Based Semantics for Agent Communication Languages. *Presented at the 1st Workshop on the History and Philosophy of Logic, Mathematics and Computation*, 2000.

[Cost et al., 2000] Cost R. S., Chen Y., Finin T. W., Labrou Y., and Peng Y.. Using Colored Petri Nets for Conversation Modeling. In *Issues in Agent Communication*, 2000, pp. 178-192.

[Dennett, 1987] Dennett D. C., *The Intentional Stance*. The MIT Press, Cambridge, MA, 1987.

[Dignum and Weigand, 1995] Dignum F., and Weigand H., Communication and Deontic Logic. In R. Wieringa and R. Feenstra, editors, *Information Systems, Correctness and Reusability*, World Scientific, Singapore, 1995, pp. 242-260.

[Dragoni, 2009] Dragoni A. F., Mental states as multi-context systems, *Annals of Mathematics and Artificial Intelligence*, Vo. 54, N. 4, 2009, pp. 265-292.

[Dragoni et al., 2002] Dragoni A. F., Giorgini P., Serafini L., Mental states recognition from communication. *Journal of Logic and Computation*, Vol. 12, N. 1, 2002, pp. 119-136.

[Dragoni and Puliti, 1994] Dragoni A. F., and Puliti P., Mental states recognition from speech acts through abduction. In *Proceedings of the 11th European Conference in Artificial Intelligence*, John Wiley & Sons, 1994.

## BIBLIOGRAFIA

[Fikes and Nilsson, 1971] Fikes R., and Nilsson N., STRIPS: A new Approach to the application of Theorem Proving to Problem Solving. In *2nd IJCAI-71*, Imperial College, London, 1971.

[FIPA] FIPA home page <http://www.fipa.org/>.

[Fisher, 1994] Fisher M., A survey of Concurrent METATEM – the language and its applications. In D. M. Gabbay and H. J. Ohlbach, editors, *Temporal Logic – Proceedings of the First International Conference*, Vol. 827, 1994, pp. 48-505.

[Flores and Kremer, 2001] Flores R.A., and Kremer R. C., Bringing Coherence to Agent Conversations. *Proceedings of the 2nd Workshop on Agent-Oriented Software Engineering. 5th International Conference on Autonomous Agent*, 2001, pp. 85-92.

[Franklin and Graesser, 1996] Franklin S., and Graesser A., Is it an Agent, or Just a Program?: A Taxonomy for Autonomous Agents. In *Proceedings of the Workshop on Intelligent Agents III, Agent Theory, Architectures, and Languages*, Springer-Verlag, 1996, pp. 21-35.

[Genesereth and Nilson, 1986] Genesereth M. R., and Nilsson N. J., *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, Inc., 1986.

[Georgeff et al., 1999] Georgeff M., Pell B., Pollack M., Tambe M., and Wooldridge M., The Belief-Desire-Intention Model of Agency. In Jörg Müller, Munindar P. Singh, and Anand S. Rao, editors, *Proceedings of the 5th International Workshop on Intelligent Agents V : Agent Theories, Architectures, and Languages (ATAL-98)*, Springer-Verlag: Heidelberg, Germany, Vol. 1555, 1999, pp. 1-10.

[Georgeff and Ingrand, 1989] Georgeff M. P., and Ingrand F. F., Decision-making in an embedded reasoning system. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, Detroit, MI, 1989, pp. 972-978.

[Gmytrasiewicz et al., 1991] Gmytrasiewicz P.J., Durfee E.H., Wehe D.K., The utility of communication in coordinating intelligent agents, formalisms for coordination. In *Proceedings of the Ninth National Conference on Artificial intelligence*, 1991.

[Gmytrasiewicz and Durfee, 1995] Gmytrasiewicz P.J., Durfee E.H., A rigorous, operational formalization of recursive modeling.. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS95)*, San Francisco, 1995, pp. 125-132.

[Greaves et al., 1999] Greaves M., Holmback H., and Bradshaw J. M., What is a Conversation Policy ?. In M. Greaves and J. M. Bradshaw, editors, *Proceedings of the Autonomous Agents '99 Workshop on Specifying and Implementing Conversation Policies*, 1999.

[Halpern et al., 1995] Halpern J. Y., Fagin R., Moses Y., and Vardi M. Y., *Reasoning about Knowledge*. MIT Press, 1995.

[Jennings and Wooldridge, 1996] Jennings N. R., and Wooldridge M., Software Agents. *IEEE Review*, 1996, pp. 17-20.

[Jennings et al., 1998] Jennings N. R., Sycara K., and Wooldridge M., A Roadmap of Agent Research and Development. *Journal of Autonomous Agents and Multi-Agent Systems*, Vol. 1, N. 1, 1998, pp. 7-38.

[Jennings et al., 2001] Jennings N. R., Faratin P., Lomuscio A. R., Parsons S., Sierra C., and Wooldridge M., Automated Negotiation: Prospects, Methods and Challenges. *Journal of Group Decision and Negotiation*, 2001, pp. 199-215.

## BIBLIOGRAFIA

[Kay, 1984] Kay A., Computer Software. *Scientific American*, Vol 251, N. 3, 1984, pp. 53-59.

[Labrou, 1996] Labrou Y., Semantics for an Agent Communication Language. PhD thesis, University of Maryland Baltimore County, 1996.

[Labrou and Finin, 1998] Labrou Y., and Finin T. W., Semantics and Conversations for an Agent Communication Language. In M. Huns and M. Singh, eds., *Readings in Agent*, 1998, pp. 235-242.

[Maes, 1997] Maes P., Agents that Reduce Work and Information Overload. In Jeffrey M. Bradshaw, editor, *Software Agents*, AAAI Press / The MIT Press, 1997, pp. 145-164.

[Mascardi and Eiter, 2002] Mascardi V., and Eiter T., Comparing Environments for developing software agents. *AI Communications*, Vol. 15, N. 4, 2002, pp. 169-197.

[Mascardi et al, 2005] Mascardi V., Demergasso D., Ancona D., Languages for Programming BDI-style Agents: an Overview. In Flavio Corradini, Flavio De Paoli, Emanuela Merelli, and Andrea Omicini, editors, *WOA*, Pitagora Editrice Bologna, 2005, pp. 9-15.

[McCarthy, 1978] McCarthy J., Ascribing mental qualities to machines. Technical Report, Stanford University AI Lab., Stanford, CA 94305, 1978.

[Mey, 2001] Mey J. L., Pragmatics. *An Introduction*. Blackwell, 2001.

[Negroponte, 1997] Negroponte N., Agents: From Direct Manipulation to Delegation. In Jeffrey M. Bradshaw, editor, *Software Agents*, AAAI Press / The MIT Press, 1997, pp. 57-66.

[Nwana, 1995] Nwana H. S., Software Agents: An Overview. *Knowledge Engineering Review*, Vol. 11, N. 3, 1995, pp. 205-244.

[Pedersen, 2001] Pedersen H. M., Agent Communication Languages - A Pre MSc Report., Oct, 2001.

[Pedersen, 2002] Pedersen H. M., Speech act and agent – A semantic analysis. Master's Thesis, Informatics and Mathematical Modelling, Thechnical University of Denmark, DTU, 2002.

[Rao, 1996] Rao A. S., AgentSpeak(L): BDI agents speak out in a logical computable language. In *Proc. of MAAMAW96*, 1996, pp. 42-55.

[Rao and Georgeff, 1991a] Rao A. S., and Georgeff M. P., Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, Sydney, Australia, 1991, pp. 498-504.

[Rao and Georgeff, 1991b] Rao A. S., and Georgeff M. P., Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR&R-91)*, Morgan Kaufmann Publishers, San Mateo, CA, 1991, pp. 473-484.

[Rao and Georgeff, 1993] Rao A. S., and Georgeff M. P., A model-theoretic approach to the verification of situated reasoning system. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, Chambéry, Francia, 1993, pp. 318-324.

## BIBLIOGRAFIA

[Rao and Georgeff, 1995] Rao A. S., and Georgeff M. P., BDI Agents: from theory to practice. In *Proceedings of the First International Conference on Multi-Agent System (ICMAS-95)*, San Francisco, CA, June 1995, pp. 312-319.

[Rao and Murray, 1994] Rao A. S., and Murray G., Multi-agent mental state recognition and its application to air-combat modeling. In *Proceedings of the Distributed Artificial intelligence Workshop*, Lake Quinalt, Washington, 1994.

[Searle, 1969] Searle J. R., *Speech Acts*. Cambridge University Press, Cambridge UK, 1969.

[Searle and Vanderveken, 1985] Searle J. R., and Vanderveken D., *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge UK., 1985.

[Shardlow, 1990] Shardlow N., Action and agency in cognitive science. Master's thesis, Department of Psychology, University of Manchester, Oxford Rd., Manchester M13 9PL, UK, 1990.

[Sharp, 2000] Sharp R., *Principles of Protocol Design*. IMM – DTU, draft second edition, 2000.

[Shoham, 1990] Shoham Y., Agent-oriented programming. Technical Report STAN-CS-1335-90, Computer Science Department, Stanford University, Stanford, CA 94305, 1990.

[Shoham, 1997] Shoham Y., An Overview of Agent-Oriented Programming. In Jeffrey M. Bradshaw, editor, *Software Agents*, AAAI Press / The MIT Press, 1997, pp. 271-290.



[Singh, 1998] Singh M. P., Agent Communication Languages: Rethinking the Principles. *IEEE Computer*, Vol. 31, 1998, pp. 40-49.

[Singh, 2000a] Singh M. P., A Social Semantics for Agent Communication Languages. *Proceedings of the IJCAI Workshop on Agent Communication Languages*, 2000.

[Singh, 2000b] Singh M. P., Commitment machines. Technical report, 2000.

[Smith et al., 1997] Smith D. C., Cypher A., and Spohrer J., KidSim: Programming Agents without a Programming Language. In Jeffrey M. Bradshaw, editor, *Software Agents*, AAAI Press / The MIT Press, 1997, pp. 167-190.

[Webster, 1913] Webster Dictionary, 1913. Disponible su: <http://machaut.uchicago.edu/cgi-bin/WEBSTER.sh?WORD=agent>.

[Weiss, 2000] Weiss G., *Multiagent Systems - A Modern Approach to Distributed Artificial Intelligence*, MIP Press, 2000.

[Winograd, 1987] Winograd T., A Language/Action Perspective on the Design of Cooperative Work. *Human-Computer Interaction*, Vol. 3, N. 1, 1987, pp. 3-30.

[Wooldridge, 1995] Wooldridge M., *Multi Agent Systems*. John Wiley & Sons Ltd, 2002.

[Wooldridge and Jennings, 1995] Wooldridge M., and Jennings N. R., Intelligent agents: Theory and practice. *The Knowledge Engineering Review* Vol. 10, N. 2, 1995, pp. 115–152.

## BIBLIOGRAFIA

[Zambonelli et al., 2003] Zambonelli F., Jennings N, and Wooldridge M., Developing Multiagent Systems: The Gaia Methodology. *ACM Transactions on Software Engineering and Methodology*, Vol. 12, N. 3, 2003, pp. 317-370.