UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

LÓGICAS BDI FUZZY

ANDERSON PAIVA CRUZ

Natal/RN
Setembro de 2008

ANDERSON PAIVA CRUZ

LÓGICAS BDI FUZZY

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Sistemas e Computação do Departamento de Informática e Matemática Aplicada da Universidade Federal do Rio Grande do Norte, como parte dos requisitos necessários para obtenção do título de Mestre.

Orientador:
Prof. Dr. Regivan Hugo Nunes Santiago
Co-orientador:
Prof. Dr. Benjamín René Callejas Bedregal

Natal/RN
2008

Catalogação da Publicação na Fonte. UFRN/Biblioteca Central Zila Mamede

Divisão de Serviços Técnicos

Cruz, Anderson Paiva
Lógicas BDI Fuzzy / Anderson Paiva Cruz.- Natal, RN, 2008.
125 p.

Orientador: Regivan Hugo Nunes Santiago.
Co-Orientador: Benjamín René Callejas Bedregal.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro de Ciências Exatas e da Terra. Departamento de Informática e Matemática Aplicada. Programa de Pós-Graduação em Sistemas e Computação.

1. Lógica (Computação) - Dissertação. 2. Lógica BDI – Dissertação. 3. Lógica Fuzzy – Dissertação. 4. Lógica modal – Dissertação. 5. Agentes inteligentes – Dissertação. 6. Ciência da computação – Dissertação. I. Santiago, Regivan Hugo Nunes. II. Bedregal, Benjamín René Callejas III.Título.

RN/UF/BCZM CDU 004.312 (043.3)

Anderson Paiva Cruz

Lógicas BDI Fuzzy

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Ciências da Computação pelo Programa de Pós-Graduação em Sistemas e Computação da Universidade Federal do Rio Grande do Norte.

Prof. Dr. Regivan Hugo Nunes Santiago

Prof. Dr. Benjamín René Callejas Bedregal

Profª. Drª. Anne Magály de Paula Canuto

Profª. Drª. Maria da Paz Nunes de Medeiros

Prof. Dr. Antônio Carlos da Rocha Costa

Agradecimentos

A Deus, por Ele tornar possível o despertar de um novo dia. Por ter me dado discernimento e sabedoria, ter me dado a tranquilidade necessária nas horas difíceis e alegria incessante. E por ter selecionado rigorosamente aqueles que estão ao meu redor.

Ao professor Dr. Regivan, meu orientador, que me aceitou e me orientou enquanto foi possível.

Ao meu tutor e eterno mestre, Professor Dr. Benjamín, que há tempos tem cuidado da minha educação e tanto me ajuda na conquista de uma boa capacitação.

Ao professor Dr. João Marcos, que me ajudou em horas fundamentais desta caminhada.

Aos professores Rocha, Anne e Maria da Paz, que aceitaram este compromisso e se dispuseram a tratar este trabalho com atenção, no intuito de me dirigir críticas sempre construtivas.

À minha mãe e ao meu pai, Professores Ângela e Maelson, os quais me iniciaram e me encorajaram a alcançar o que quer que eu almejasse. Pela sua paciência e compreensão.

À minha irmã, Michelle, que me fazia companhia durante os momentos de estudo e por me incentivar a trabalhar mais arduamente neste projeto.

Aos amigos do mestrado, da igreja e da família, que foram providenciais em momentos paraconsistentemente sérios e descontraídos.

Ao CNPq por ter financiado este trabalho de pesquisa. E aos funcionários e professores que compõem o DIMAp.

Por fim, aos que me ajudaram de alguma forma, mas que por impossibilidade de lembrar de todos, não foram mencionados.

Resumo

Com o intuito de entender como a mente humana funciona iniciaram-se estudos sobre cognição nos campos da filosofia e psicologia. Teorias surgiram desses estudos e, atualmente, esta curiosidade foi estendida a outras áreas, tais como, ciência e engenharia de computação, no entanto, nestas áreas, o objetivo é sutilmente diferente: entender o funcionamento da mente e aplicá-lo em uma modelagem artificial. Em ciência da computação, a sub-área de sistemas multiagentes tem progredido bastante, utilizando trabalhos em inteligência artificial, lógica computacional, sistemas distribuídos, teoria dos jogos e, aproveitando também teorias provenientes da própria filosofia e psicologia. Desta forma, alguns pesquisadores já vêem o paradigma de programação orientado a agentes como a melhor solução para a implementação dos *softwares* mais complexos: cujos sistemas são dinâmicos, não-determinísticos e que podem ter de operar com dados faltosos sobre ambientes também dinâmicos e não-determinísticos. Este trabalho busca a apresentação de uma extensão da formalização lógica de um modelo de arquitetura de agentes cognitivos, chamado BDI (*belief-desire-intention*), na qual o agente é capaz de deliberar suas ações baseando-se em suas crenças, desejos e intenções. A formalização de tal modelo é conhecida pelo nome de lógica BDI, uma lógica modal com três relações de modalidade. Neste trabalho, serão apresentados dois planos para transformá-la numa lógica modal fuzzy onde as relações de acessibilidade e as fórmulas (modais-fuzzy) poderão ter valorações dentro do intervalo [0,1]. Esta lógica modal fuzzy há de ser um sistema lógico formal capaz de representar quantitativamente os diferentes graus de crenças, desejos e intenções objetivando a construção de raciocínios fuzzy e a deliberação de ações de um agente (ou grupo de agentes), através dessas atitudes mentais (seguindo assim um modelo intensional).

Abstract

Intendding to understand how the human mind operates, some philosophers and psycologists began to study about rationality. Theories were built from those studies and nowadays that interest have been extended to many other areas such as computing engineering and computing science, but with a minimal distinction at its goal: to understand the mind operational proccess and apply it on agents modelling to become possible the implementation (of softwares or hardwares) with the agent-oriented paradigm where agents are able to deliberate their own plans of actions. In computing science, the sub-area of multiagents systems has progressed using several works concerning artificial intelligence, computational logic, distributed systems, games theory and even philosophy and psycology. This present work hopes to show how it can be get a logical formalisation extention of a rational agents architecture model called BDI (based in a philosophic Bratman's Theory) in which agents are capable to deliberate actions from its beliefs, desires and intentions. The formalisation of this model is called BDI logic and it is a modal logic (in general it is a branching time logic) with three access relations: B, D and I. And here, it will show two possible extentions that tranform BDI logic in a modal-fuzzy logic where the formulae and the access relations can be evaluated by values from the interval [0,1].

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Apresentar-se-á aqui uma nova maneira de modelar agentes cognitivos. Eles são fontes de pesquisa em diversos campos da ciência: nas áreas de economia (principalmente ao que se refere a sub-área de micro-economia e sua íntima relação com a teoria dos jogos – como por exemplo [NM44]), filosofia (entre vários autores tais como Searle, J. L. Austin, Steven Pinker, George Lakoff, D. Vanderveken, etc., pode-se citar, ainda, contribuições de Dennet, como por exemplo [Den87] e [Den97]), psicologia [Sti83] e ciência da computação, principalmente nas sub-áreas de inteligência artificial ([MW99] e [Gär88]), sistemas multiagentes ([Woo02a, LCN90, Jer92]) e lógica computacional ([Woo00a, RG93, RG98, CL90, vLvdHM98]).

Um agente é qualquer entidade que interage com o ambiente (seja este físico ou lógico – software), onde o mesmo está inserido. O comportamento de tal entidade pode ser definido simplificadamente em função dos estados do ambiente e as ações às quais ele realiza (vide figura 1 retirada de [Woo02a]). Daí o agente percebe como está o ambiente (efetua percepção), transforma-o em informação e resolve qual ação deve ser tomada. Esta, por sua vez, irá modificar o ambiente e o agente deverá novamente obter uma nova percepção sobre o ambiente. Surge, então, o seguinte questionamento: Como o agente decide? Ou melhor, quais mecanismos internos fazem com que ele escolha, uma determinada ação a partir da sua percepção sobre o ambiente? A resposta para tal pergunta dependerá da definição da **arquitetura interna do agente**.

Figura 1.1: Comportamento do(s) agente(s) com relação ao ambiente.

Há atualmente três tipos de arquitetura cognitiva: A **reativa**, na qual o agente age exclusivamente a partir da percepção que ele tem sobre o ambiente; a **cognitiva**, na qual o agente é capaz de raciocinar sobre o ambiente ou seu histórico, planejando e deliberando suas ações futuras; e a **híbrida**, em que o agente é hábil para agir deliberada ou reativamente (de acordo com Wooldridge em [Woo02a]). Há diversas formas de implementar computacionalmente estes tipos de arquitetura. E não há um tipo melhor que o outro, pois depende de que tipo de mente querer-se-á modelar. Por exemplo: uma lagarta pode-se supor que ela atue apenas por instinto comendo sempre as folhas sem que, para ela, haja qualquer objetivo aparente; uma leoa, pode agir por instinto ao defender seus filhotes de um outro animal mais forte ou pode planejar um método de capturar sua presa, pois ele precisa alimentar-se e a seus filhotes; e um humano que, quando sábio, age deliberadamente programando toda a sua vida e suas ações.

Wooldridge e Jernings, em [WJ95, Woo00a], afirmam que o agente (cognitivo) possui quatro propriedades:

- **Autonomia:** agentes devem agir sem a intervenção direta de outros agentes e devem ter algum controle sobre suas ações e estados internos.
- **Reatividade:** agentes percebem o ambiente e respondem temporalmente a mudanças no mesmo, ou seja, eles interagem com o ambiente.
- **Pró-atividade:** agentes não agem simplesmente em resposta ao ambiente em que se encontram, eles são aptos a tomar decisões de acordo com o seu comportamento.
- **Habilidade social:** agentes interagem entre si via alguma linguagem de comunicação de agentes, não apenas trocando dados, mas também convencendo ou negociando para requerer ou obter alguma ação ou estado mental.

Uma das formas de se implementar uma arquitetura cognitiva e, desta forma, modelar um agente cognitivo, é através de uma arquitetura baseada na teoria BDI. Esta teoria se baseia em três atitudes mentais para deliberar quais ações serão executadas pelo agente, são elas: crença (*Belief*), desejo (*Desire*) e intenção (*Intention*). Neste trabalho, tentar-se-á mostrar duas maneiras de estender uma formalização lógica dessa teoria com o intuito de aproximá-la ainda mais da realidade, ou melhor, tornar o raciocínio do agente inteligente ainda mais fidedígno à sua natureza, através da ultilização da lógica fuzzy. Construindo, assim, lógicas BDI fuzzy a qual é apropriada para modelar o raciocínio aproximado em situações onde há incertezas.

1.1 Motivação

O porquê da formalização lógica de tal teoria é a primeira motivação encontrada para concretização deste trabalho. A teoria BDI é bastante discutida no âmbito filosófico; enquanto o modelo BDI é muito aplicado em algoritmos da área de inteligência artificial e sistemas multiagentes. Entretanto, nenhuma dessas duas classes de estudos se preocupa em analisar as relações e prioridades das atitudes mentais de forma explícita e sem a ambiguidade inerente à linguagem natural. A lógica BDI é uma lógica modal temporal de ações que possibilita provar (para entender melhor) propriedades com respeito as intenções, desejos e crenças. Assim, a relação de acessibilidade B gerada a partir do conectivo modal Bel determina para cada mundo (estado do ambiente) $w \in W$ – onde W é um conjunto de mundos –, quais mundos de W são compatíveis com as crenças de um dado agente i no mundo w. Ou seja, dado um conjunto de crenças Cre do agente i no mundo w, um mundo $w' \in W$ é acessível a partir de w se, e somente se, (sse para simplificar) toda crença $\varphi \in Cre$ for verdadeira em w'. Dessa forma, Bel_i^w determina o conjunto de estados possíveis do ambiente, de acordo com as crenças do agente i a respeito do ambiente em que se encontra (mundo w). Analogamente ocorre para a relação de acessibilidade D. Assim, Des_i^w determina o conjunto de estados desejáveis do ambiente, conforme os objetivos de i em relação ao ambiente; e Int_i^w determina o conjunto de estados do ambiente os quais i está efetivamente comprometido em alcançar.

Ao estender a lógica BDI no sentido de uma Lógica Modal Fuzzy, estar-se-á introduzindo graus de pertinência às relações e às sentenças envolvidas na semântica de Kripke de modo a obter conectivos modais ou relações de acessibilidade fuzzy para representar algumas funções mentais de agentes que possuem graus variáveis de crenças, desejos e intenções, possibilitando a modelagem de agentes capazes de ter dúvidas e vontades distintas em relação a objetivos distintos (ou até mesmo idênticos em instantes de tempo diferentes). O que o torna mais próximo da natureza mental dos agentes cognitivos que mais conhecemos (os seres humanos); por exemplo: não se pode obrigar que um homem acredite fielmente em tudo, tem que se permitir a possibilidade de dúvidas ao mesmo! Da mesma forma que se deve possibilitar que um agente queira conseguir mais uma coisa do que outra, como por exemplo, um agente pode desejar mais ser poligâmico do que não ser criticado pela sua sociedade monogâmica e assim decidir no que ele irá se comprometer; ou ainda, se o agente não tiver certeza da cor exata do seu computador (supostamente cinza) e por isso atribui os seguintes valores verdade às sentenças do seu sistema: Bel_i^wCOR_DO_COMPUTADOR$(preto) = 0,5$ e Bel_i^wCOR_DO_COMPUTADOR$(branco) = 0,5$ sem provocar a trivialização do mesmo.

Pode-se citar ainda inúmeras abstrações, entre elas: suponha que um agente quer saborear uma pizza e um sorvete, mas só possui dinheiro para um deles, logo ele opta por resolver parcialmente seu plano escolhendo comprar a refeição que ele tem a maior vontade naquele instante. Ou caso ele deseje se tornar um engenheiro e ao mesmo tempo queira, com um grau menor, se tornar apicultor; baseando-se nisso, ele pode escolher opções de planos que o torne um engenheiro, mas que não descarte a possibilidade de se tornar também um apicultor, como ler livros sobre o assunto ou fazer um curso de pequena duração ou simplesmente comprar um criadouro de abelhas. [1] Perceba que tais exemplos influem diretamente na deliberação dos planos do agente e intuem a possibilidade de usar sistemas de inferência fuzzy na modelagem de sistemas orientados a agentes – o que até então não foi conquistado por nenhuma formalização lógica BDI. Bem como, possibilitam a resolução de um dos problemas críticos da especificação de agentes BDI que é o chamado problema da desistência de uma intenção (quais critérios o agente deve usar para desistir ou não de um objetivo?).

Além da aplicação de raciocínio aproximado nos estados mentais do agente, proporcionado pela fuzzyficação dos mesmos, é possível também fuzzyficar o ambiente (o mundo) em que ele se encontra através da definição de uma semântica fuzzy à interpretação de conectivos temporais e daqueles provenientes da lógica clássica ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$). Aproximando, desta forma, o agente de um agente com características mais humanas,

[1] Tais abstrações caso fossem operados sobre um sistema BDI clássico, o mesmo teria que extinguir uma das intenções utilizando algum motivo obscuro ao agente (aleatoriamente ou segundo uma fila de intenções).

e aproximando o ambiente do contexto em que vivemos.

Outra vantagem na criação de um sistema de inferência fuzzy BDI é solucionar alguns aspectos do problema da resolução de planos parciais, pois podem ser usados $\alpha-$cut nos conjuntos de desejos/intenções provocando a eliminação de desejos/intenções cujo grau de verdade são menores do que α e assim determinar o que será executado agora e o que será executado posteriormente.

Este trabalho, portanto, propõe mostrar formalizações lógicas BDI, baseada na teoria de raciocínio prático desenvolvido por Bratman ([Bra87]), capaz de representar um sistema multiagente composto por agentes BDI cujas crenças, desejos e intenções são quantificadas em graus do intervalo [0,1]. E assim, o agente inteligente não precisa ser isento de dúvidas – pois suas crenças possuem valorações para determinar o quanto ele acredita nelas – bem como, ele não será mais obrigado a ter a mesma vontade de realizar todos os seus desejos e intenções; como feito nas formalizações BDI até então.

Essas novas formalizações serão realizadas através da extensão de uma lógica BDI definida neste mesmo trabalho tornando-a uma lógica BDI fuzzy. A partir desta extensão, poder-se-á definir várias lógicas (semânticas) em que é possível valorar ou não as relações de acessibilidade e as relações entre estados temporais.

1.2 Modelagem de Agentes Cognitivos

Os agentes cognitivos podem ser modelados com um raciocínio teórico ou com raciocínio prático. O raciocínio teórico é o raciocínio orientado a crenças e é um processo de conclusão, ou melhor, de dedução natural (em geral suas implementações utilizam provadores automáticos de teoremas). Enquanto o raciocínio prático é um raciocínio orientado a ações e consiste de dois processos:

1. Deliberação de quais estados devem ser atingidos.
2. Raciocínio "meio-fim" (ou "*means-ends*") que determina como chegar a esses estados (quais ações devem ser escolhidas para se chegar a tal resultado).

Um exemplo simples para diferenciar esses dois tipos de raciocínio é: o agente conclui (ou deduz) que irá chover a partir de suas crenças de que o céu está nublado e o tempo está esfriando, é um raciocínio teórico; e o raciocínio prático seria o raciocínio do que o agente deve fazer para não se molhar. Aqui, concentrar-se-á em modelagem de agentes com raciocínio prático.

1.2.1 Modelagem de Agentes Cognitivos com Raciocínio Prático

Segundo Wooldridge [Woo02a]: "Agentes mantém uma representação interna do seu mundo, tal que existe um estado mental que pode ser modificado por alguma forma de raciocínio simbólico" (tradução livre).

Tal estado mental é uma "atitude" (mental) escolhida para caracterizar o agente e Hoek e Wooldridge em [HW03], elas são subdivididas da seguinte forma:

Atitudes ligadas às Informações: conhecimento e crença;

Pró-Atitudes: objetivos, desejos e intenções;

Atitudes Normativas: autorizações, permissões e obrigações.[2]

A representação interna ou estado interno do agente é o conhecimento que o mesmo tem sobre ele próprio e sobre o ambiente em que ele se encontra. Assim, pode-se pensar no funcionamento de um agente cognitivo da seguinte forma (retirado de [Woo02a]):

A partir de um estado inicial C^0 (representação simbólica do seu conhecimento), o agente observa o ambiente A^0 e obtém uma percepção P^0 sobre o mesmo (representação simbólica externa) o que implica em C^1 ($C^1 := C^0 + P^0$). Em função de C^1, o agente delibera uma ação a_1 que, por sua vez, poderá mudar o ambiente para A^1.

Entretanto, surge a seguinte questão: Como se procede a transformação de percepção em conhecimento e de conhecimento em ação? Existem inúmeras maneiras de responder tal pergunta e aqui serão abordadas três delas: uma, segundo a vertente da lógica de Cohen e Levesque; uma, baseada nos resultados e oportunidades do agente [vLvdHM98]; e outra, fundamentada nas crenças, desejos e intenções (Modelo BDI).

Por fim, é importante explicitar que as formalizações a serem apresentadas limitam-se em analisar o agente como uma entidade isenta de emoções, ou, pelo menos de uma forma em que estas não são consideradas diretamente. Visualiza-se um agente apenas do ponto de vista de uma postura intensional. A postura intensional, segundo Dennet em [Den97], é o meio pelo qual se interpreta o comportamento de uma entidade considerando-a como um agente inteligente capaz de fazer escolhas de ações através das suas crenças e desejos.

1.3 Organização do Trabalho

O segundo capítulo será dedicado a tratar da modelagem de agentes cognitivos; primeiramente, falando das propriedades requeridas pelos mesmos e, em seguida, tratando das três principais (mais conhecidas) formalizações, segundo [HW03], de arquiteturas internas de agentes cognitivos: a lógica de Cohen e Levesque, "*KARO*" cujo processo de decisão envolve o conhecimento do agente (*Knowledge*), suas ações (*Action*), resultados (*Results*) e oportunidades (*Opportunities*) e, finalmente, a teoria BDI dando um enfoque a uma lógica BDI definida no sexto capítulo deste trabalho a qual tem como base a lógica BDI de Rao e Georgeff ([RG91, RG93]) e a LORA (*Logic of Rational Agents*) de Wooldridge.

Nos dois capítulos seguintes serão vistas as duas lógicas que serão utilizadas para construir a lógica modal fuzzy, objetivo da dissertação: no terceiro capítulo, será apresentada a lógica modal clássica e a lógica temporal, no quarto, apresentar-se-á a lógica fuzzy de Zadeh desenvolvida a partir da Teoria de conjuntos Fuzzy ([Zad65]); no quinto capítulo, serão apresentadas com maiores detalhes as lógica BDI de Rao e Georgeff e de

[2][HW03] afirma que não há um consenso de quais atitudes devem ser usadas, por isso as teorias de modelagem de agentes apenas as escolhem intuitivamente e estudam as possíveis relações entre elas.

Wooldridge[3] e se explicará porque elas não serão adotadas nas propostas de lógicas BDI fuzzy definidas no sétimo capítulo, as quais são baseadas na lógica BDI definida no sexto capítulo.

Por fim, no oitavo capítulo é estudado um caso de sistema multiagente, onde há a aplicação das formalizações BDI e BDI fuzzy para analisar o raciocínio de um dos agentes. E o nono capítulo, tratar-se-á da conclusão acerca dos resultados deste trabalho.

[3]Tais lógicas são as únicas lógicas BDI da literatura cujos autores assumiram seguir unicamente a teoria de intenções de Bratman

Capítulo 2

Agentes Cognitivos

Antes de falar sobre os agentes BDI, para um melhor entendimento, é interessante detalhar mais sobre agentes racionais (também chamados de cognitivos ou inteligentes) e sobre conceitos relacionados a este tema.

Agentes cognitivos, segundo a definição de cognição de [Fer86], são aqueles agentes que possuem a capacidade de adquirir conhecimento, aprender, o que facilmente remete à noção de que o agente é capaz de raciocinar antes de agir, decidir depois de examinar, enfim: deliberar. Dessa forma, é inerente ao agente cognitivo a capacidade de armazenar o conhecimento ou, pelo menos, conseguir representá-lo simbolicamente. Daí, a primeira importância da formalização da arquitetura interna deste tipo de agente. Além disso, através da formalização, há um melhor entendimento e uma investigação mais segura das suas propriedades, livre das ambiguidades da linguagem natural.

2.1 Formalização de Cohen e Levesque

A formalização de Cohen e Levesque, segundo [HW03], foi usada inicialmente para desenvolver uma teoria de intenção (será detalhada no final deste capítulo), pois tal teoria, segundo os próprios autores da formalização, é um pré-requisito para desenvolver a Teoria de Atos da Fala a qual pode ser melhor estudada em [Woo02b, Woo00b], no entanto, esta lógica vem comprovando o seu bom uso na especificação de raciocínio sobre as propriedades de agentes com problemas de cooperação em sistemas multiagentes ([Gal88a, Gal88b, Jer92, LCN90]).

Em [LCN90] é especificado um balanço racional ("*rational balance*") necessário entre crenças, objetivos, planos, intenções e compromissos. Esse balanceamento se mostra necessário, pois, se um agente possui uma dada intenção, então, o mesmo precisa descrever como ela afeta suas crenças e como se comprometer com as ações futuras. Tal formalização segue duas fundamentações filosóficas de teorias de intenção: Bratman ([Bra87, Bra84]) e Searle ([Sea83]). Em [Bra87], Bratman identifica sete propriedades as quais devem ser satisfeitas em uma teoria de intenção:

1. Intenções apresentam problemas os quais os agentes devem determinar um modo de resolvê-los,

2. Intenções fornecem um filtro para que outras intenções não conflitantes possam ser adotadas,

3. Agentes tentam ser bem sucedidos em suas intenções, mas se falharem são inclinados a tentarem novamente,

4. Agentes acreditam que suas intenções são exequíveis,

5. Agentes não acreditam que não conseguirão realizar suas intenções,

6. Sob certas circunstâncias, os agentes acreditam que eles conseguirão realizar suas intenções,

7. Agentes não precisam pretender os efeitos causados pela realização de suas intenções.[1]

A partir desses critérios (baseando-se principalmente nos itens 5 e 7), os autores construíram uma formalização de agentes racionais multi-tipos, multi-modal com equivalências onde a intenção é definida como um conceito composto especificando o que o agente escolheu e como ele irá se comprometer com tal escolha. A formalização é desenvolvida em duas camadas. Na primeira, encontra-se as modalidades primárias (tabela 2.1), operadores sobre as ações (tabela 2.2) e análises acerca de crenças, desejos e ações. Na segunda, se desenvolve uma "teoria parcial de ação racional", definindo a intenção como um objetivo persistente $P - Goal$ (equação 2.1 a qual é detalhada posteriormente) e os operadores derivados explicitados na tabela 2.3.

Operador	**Significado**
$BEL_i\varphi$	agente i acredita em φ
$GOAL_i\varphi$	agente i tem o objetivo em φ
$HAPPENS\,\alpha$	ação α acontecerá em seguida
$DONE\,\alpha$	ação α acabou de acontecer

Tabela 2.1: Modalidades primárias da Formalização de Cohen e Levesque

Observação 2.1.1 *φ e α, nesta e em todas as outras formalizações apresentadas neste trabalho, são proposições. A primeira pode representar uma propriedade ou um estado mental; enquanto a segunda representa uma tarefa a ser realizada.*

Operador	**Significado**
$\alpha; \alpha'$	α seguindo de α'
$\varphi?$	teste de que φ ocorre

Tabela 2.2: Operadores sobre ações da Formalização de Cohen e Levesque

$$P - Goal_i(p) \stackrel{\text{def}}{=} Goal_i(LATERp) \wedge Bel_i\neg p \wedge$$
$$Before((Bel_i p) \vee (Bel_i \Box\neg p))(\neg(Goal_i(LATERp)) \tag{2.1}$$

Dessa forma um agente possui um objetivo persistente p se:

[1]Por exemplo: um agente tem a intenção de comer doces após o almoço, sabe que isso o deixará mais gordo, mas ele pode não ter a intenção de se tornar mais gordo.

Operador	**Significado**
$\Diamond\varphi$	Há uma sequência de eventos em que depois dela φ se torna verdadeira
$\Box\varphi$	φ é sempre verdadeiro
$BEFORE\,p\,q$	a proposição p é verdadeira antes da proposição q
$LATER\,p$	p não é verdadeiro agora mas é depois
$KNOW\,i\,p$	o agente i sabe que p
$COMPETENT\,i\,p$	o agente i é competente com respeito à proposição p, ou seja, ele só adota crenças em relação à p se ele tiver boas evidências de p

Tabela 2.3: Modalidades Secundárias da Formalização de Cohen e Levesque

- O agente tem o objetivo de que p eventualmente se torne verdadeiro e acredita que p, agora, não é verdadeiro;
- O objetivo só é descartado se o agente acredita que o objetivo foi satisfeito ou se ele acreditar que não é possível satisfazê-lo.

A partir de tal definição de intenção, Cohen e Levesque afirmam que sua formalização evita que os agentes se comprometam com uma intenção já realizada – “sobreposição de compromissos idênticos” (“*overcommitment*”) – e com uma intenção inexequível – “falso-compromisso” (“*undercommitment*”).

Semântica de Kripke As interpretações das sentenças são dadas através da semântica de Kripke, ou seja, modelos $M = \langle \Theta, P, E, Agt, T, B, G, \Phi \rangle$ onde Θ é um conjunto, P um conjunto de pessoas, E é um conjunto de eventos primitivos, Agt especifica o agente de um evento ($Agt \in [E \rightarrow P]$), T é um conjunto não vazio de mundos, B e G são duas relações (entre mundos) de acessibilidade originadas dos conectivos BEL e $GOAL$ respectivamente, tal que B é transitiva, euclidiana e serial, e assim obedece ao sistema KD45 e G é serial (sistema KD)[2]. E Φ interpreta predicados.

Os mundos são sequências lineares e discretas de eventos no passado e no futuro, isto é, em cada mundo há uma quantidade infinita de estados à frente e atrás do estado corrente e cada evento ocorre numa relação linear entre esses estados de modo que o arco entre dois estados adjacentes possa ser interpretado como a passagem de tempo. Assume-se nesta formalização a propriedade (chamada realismo) $GOAL\,i\,\varphi \rightarrow BEL\,i\,\varphi$ a qual semanticamente indicaria que um agente não pode ter um objetivo o qual ele crê que não possa ser executado.

Cohen e Levesque, em [CL90], definem ainda dois operadores de intenção $INTEND_1$ e $INTEND_2$ para definir outros tipos de compromisso de intenções para ações e para estados mentais (como por exemplo “ser rico” ou “ser feliz”).

[2]Esses sistemas modais serão detalhados no terceiro capítulo.

2.2 Estrutura KARO

Em [vLvdHM98], Linder, Hock e Meyer apresentam *KARO FRAMEWORK*: uma formalização lógica proposicional multi-modal, dinâmica e exógena (as atitudes de habilidade e oportunidade são tratadas explicitamente) de agentes racionais. Ela trata o comportamento do agente a paritr da sua lógica de ação. Por isso, a estrutura KARO é baseada nas atitudes de conhecimento ("*Knowledge*"), habilidade ("Abilities"), resultados ("Results") e oportunidades ("Opportunities") e é fundamentada nas idéias de Bob Moore contidas em [Moo77].

Linder, Hock e Meyer ([vLvdHM98]) apresentam definições formais de dois sistemas[3] pertencentes à estrutra KARO. Tais sistemas se diferem devido às duas possibilidades de interpretações de habilidades em relação à sequência de ações. Por fim, após apresentarem várias propriedades em relação às atitudes mentais, os autores elaboram dois sistemas de prova, associados às duas interpretações, corretos e completos.

Argumenta-se em tal artigo que as habilidades e oportunidades se relacionam da seguinte forma: a habilidade é uma capacidade interna física, mental ou moral do agente e a oportunidade é uma possibilidade em virtude das circunstâncias (independente do agente); assim, as habilidades podem ser exercidas apenas quando há as oportunidades para realizá-las, ou seja, a combinação de habilidade e oportunidade determina quando um agente tem a possibilidade prática[4] para realizar uma ação específica.

2.2.1 Sintaxe e Semântica KARO

A linguagem KARO é composta por um conjunto de nomes de agentes, de variáveis proposicionais e de ações atômicas, bem como, pelos operadores $K\_$ (conhecimento), $\langle do\_(\_)\_\rangle$ (oportunidade), $A\_\_$ (habilidade), $confirm\_$ (confirmações), $\_;\_$ (composição sequencial), $if\_then\_else\_$ (composição condicional) e $while\_do\_od$ (composição repetitiva). Há ainda os operadores derivados (tabela 2.4), onde $[do_i(\alpha)]\varphi$ é definido como o dual de $\langle do_i(\alpha)\rangle\varphi$, sendo este o compromisso de i realizar α e obter φ como um dos resultados desta performance.

Definição do Operador	Significado
$M_i\varphi \stackrel{\text{def}}{=} \neg K_i\neg\varphi$	i considera que φ é possível (possibilidade epistêmica)
$[do_i(\alpha)]\varphi \stackrel{\text{def}}{=} \neg\langle do_i(\alpha)\rangle\neg\varphi$	i não se compromete em realizar α mas se houver oportunidade de fazer α, φ será um dos resultados
$skip \stackrel{\text{def}}{=} confirm\top$	mantém a ação
$fail \stackrel{\text{def}}{=} confirm\bot$	aborta a ação
$\alpha^0 \stackrel{\text{def}}{=} skip$	evita a ação
$\alpha^{n+1} \stackrel{\text{def}}{=} \alpha;\alpha^n$	ação α é seguida da ação α^n

Tabela 2.4: Operadores derivados

[3]São dois sistemas em virtude das duas possíveis interpretações para as fórmulas da linguagem KARO (ou seja, a linguagem é a mesma para os duas semânticas)

[4]A noção de possibilidade prática é também detalhada em [vLvdHM98].

A interpretação também é baseada num modelo de mundos possíveis que contém um conjunto S de mundos possíveis, uma valoração π para as variáveis proposicionais, um conjunto de agentes A, um conjunto de ações Ac e as funções:

$$r_0 : A \times Ac \rightarrow S \rightarrow S$$

que indica a transição de estados causada pela performance de uma ação atômica; e

$$c_0 : A \times Ac \rightarrow S \rightarrow \{0, 1\}$$

determina as habilidades de um agente realizar ou não uma dada ação atômica.

2.3 Teoria BDI

BDI é uma teoria que se utiliza de três atitudes mentais para caracterizar e definir o comportamento de um ou mais agentes racionais. A vantagem desta teoria é possuir três componentes distintos:

Fundamentação filosófica: O modelo BDI é baseado numa teoria de ações racionais relativa a seres humanos desenvolvida por Bratman em [Bra87]. Foi ele, o primeiro a definir o comportamento de um agente a partir de três atitudes mentais (crença, desejo e intenção) e a partir delas determinar os planos a serem possívelmente adaptados.

Formalização lógica: Há uma família de lógicas, capazes de capturar as características do modelo BDI tal como [RG91, RG93] e LORA, uma lógica BDI genérica apresentada em [Woo00a], bem como, inúmeros trabalhos envolvendo procedimentos de decisões ([RG98] que apresentam um pseudo-tableaux para um tipo particular de lógicas BDI e [Rao96b] o qual apresenta trabalhos sobre teorias envolvendo agentes, arquiteturas, linguagem e modelagem de agentes).

Arquitetura de software: Este componente é a aplicação dessas duas fundamentações (ou pelo menos da primeira delas, já que as formalizações lógicas ainda possuem algumas ambiguidades, as quais serão re-avaliadas nesta dissertação) na modelagem de um agente cognitivo orientado a ações, no entanto, o modelo de agentes BDI não prescreve uma implementação específica. Em verdade, o primeiro sistema implementado foi IRMA ([BIP88]) (*"Intelligent Resource-bounded Machine Architecture"*) o qual tem a intenção de ser uma aplicação direta de [Bra87]. Mas, a implementação mais conhecida é a PRS ("*Procedural Reasoning System*"). Segundo [PL87], PRS possui como base de sua implementação uma "biblioteca de planos". Nesta biblioteca, encontram-se um conjunto de planos, chamados desejos, e a invocação de um plano depende da circunstância em que o agente está. Ou seja, o controle de chamada de um determinado plano é feito através da atualização das crenças do próprio agente. E as intenções do agente constituem a pilha de execução de ações que contém os desejos escolhidos. Há ainda "dMARS" ([dKLW97]), a linguagem AgentSpeak ([Rao96a]), a ferramenta "Jason" para trabalhar com tal linguagem ([BHW07]), entre outras.

2.3.1 Sistemas e Ambientes

Os sistemas e ambientes aos quais BDI pode ser aplicado são aqueles que necessitam de uma arquitetura orientada a agentes (cognitivos especialmente). Onde o paradigma de programação orientado a agentes – forma pela qual a programação de um hardware ou de um software utiliza concepções de agentes – é mais cabível.

Normalmente, são sistemas dinâmicos e não-determinísticos, com alto nível de gerenciamento e controle de tarefas em ambientes também dinâmicos e não-determinísticos. São esses sistemas complexos, com pouca quantidade de informação (incompletude dos dados) e ambientes onde há incerteza do que irá acontecer[5]. Em tais sistemas, apesar da possível utilização do paradigma orientado a objetos em sua implementação, este se torna uma solução custosa em função da grande quantidade de dados a serem armazenados e cálculos a serem executados; pois são características desses sistemas:

- Diferentes ações podem ser executadas e diferentes acontecimentos imprevisíveis podem ocorrer a qualquer momento (já que tanto o sistema como o ambiente são não-determinísticos e dinâmicos).
- Há vários objetivos diferentes a serem alcançados. Não importando se são conflitantes ou não.
- O resultado das ações executadas dependem dos estados do ambiente, ou seja, as escolhas das ações irão depender das circunstâncias em que o ambiente se encontra no tempo corrente.
- A taxa de procedimentos computados deve ser limitada pela taxa da frequência de mudanças do ambiente, pois não é interessante que o ambiente mude antes que um planejamento ou a realização de algum procedimento seja efetuado, uma vez que isto poderia acarretar a não coerência entre o plano, a intenção e as crenças do agente.

Como exemplo desses sistemas cita-se: OASIS (controlador de tráfego aéreo apresentado em [GPP$^+$]), E-commerce, sistemas concorrentes ou distribuidos, simuladores de combate aéreo, etc..

2.3.2 Conceitos

BDI é uma teoria usada na modelagem de agentes racionais que considera a primazia de 3 atitudes mentais essenciais à escolha de realizações de ações. São elas, *Beliefs* (crenças), *Desires* (Desejos) e *Intentions* (Intenções).

Crenças

As crenças são informações que o agente tem sobre o ambiente ou sobre o seu passado (experiência do agente). Estas informações podem estar incompletas ou incorretas devido a algum problema do responsável por adquirir os dados, seja ele um sensor (que pode ser de baixa resolução e não consegue reconhecer/definir bem os objetos

[5] Esses são os sistemas que estão crescendo cada vez mais no mercado segundo Georgeff em [GPP$^+$]

ou o limite dos mesmos, por exemplo) ou um minerador de dados mal implementado que deve obter as crenças do agente a partir de um banco de dados ou até mesmo um humano com suas limitações.

Em geral, nas lógicas epistêmicas é feita a diferenciação entre conhecimentos (K) e crenças, como já foi visto na definição de K, na formalização de Cohen e Levesque.

Desejos

Desejos são estados os quais o agente gostaria de alcançar. Dentro do contexto de um mundo ideal, todos os desejos seriam consistentes, mas, no mundo real, os agentes devem lidar com a vontade de realização de ações conflitantes ou até mesmo contraditórias, por exemplo: alguém acredita que deve ir à missa todos os domingos, mas também gostaria muito de assistir ao jogo do Alecrim Futebol Clube na televisão que para infelicidade do torcedor ocorrerá no mesmo horário da missa. Neste exemplo, o agente possui dois desejos conflitantes (ir à missa e assistir ao jogo) e a escolha de qual ação será realizada dependerá de sua intenção (com qual desejo ele irá se comprometer?) e também de suas crenças (pois o agente crê que não irá mais para o céu se faltar a missa por tal razão, por exemplo). Tais nuances serão melhor discutidas nas subseções posteriores.

Intenções

Intenções são desejos os quais o agente se compromete em realizar. Podem ser caracterizados por ações ou estados mentais, por exemplo: "Eu pretendo ir ao aniversário do meu orientador que está acontecendo agora (a intenção aqui é caracterizada por uma ação – a intenção de me dirigir à festa – e foi deliberada para ser realizada no tempo "agora" e por isso ela é dirigida ao presente e chamada de intenção presente ou "*present-directed intention*"); outro caso seria se o aniversário dele já tivesse passado e ele estivesse em viagem e por isso tenho a intenção de ir no próximo ano a sua festa de aniversário (a intenção aqui caracteriza meu estado mental). Neste exemplo, a intenção é chamada de intenção futura ou *"future-directed intention"*. Outra diferença apresentada por Bratman em [Bra87] é entre planejar e pretender; segundo o filósofo, o primeiro requer o segundo, mas o contrário não é necessário. E quanto mais as ações são influenciadas por deliberações feitas anteriormente (são parte de um plano) mais elas são impactantes ao agente (no exemplo anterior – no caso da intenção futura – eu deveria planejar me arrumar ou até comprar uma roupa nova, não marcar nenhum compromisso no mesmo dia da festa, não beber demais no dia anterior. . .).

Outra importante característica é que intenções influenciam as ações mais do que outras pró-atitudes (desejos). Por isso, se diz que elas fazem parte de um objetivo maior (mais importante). No caso hipotético da sub-seção anterior, caso a intenção fosse ir à missa, então, ações direcionadas a esta intenção deveriam ser tomadas apesar da existência de outra pro-atitude (segundo um raciocínio clássico[6]). Assim, afirma-se que as intenções dirigem o raciocínio meio-fim. Ou seja, formada uma intenção, tentar-se-á realizá-la, envolvendo entre outras coisas, decidir como realizá-la (caso a mesma seja uma intenção futura).

Entre as intenções e as crenças há algumas relações. E estas relações podem gerar inconsistências

[6]O uso desta palavra foi feito no sentido usual lógico de obediência dos princípios aristotélicos

ou incompletudes. Em [Bra87, 37-38] são apresentados exemplos de sua tese de assimetria[7], a qual diz que ter intenção de α não requer acreditar que α será realizado (em função do agente não acreditar em sua competência em executar α ou pelo ambiente não propiciar a realização de α). Por isso, Bratman não requer a suposição forte de que para pretender α o agente deve crer que α ocorrerá. Ou seja, é normal a intenção em algo que se acredite que ocorrerá, no entanto, não é irracional pretender algo que não se creia que ocorrerá (há neste caso apenas uma incompletude entre intenções e crenças). Entretanto, seria irracional (inconscistente) pretender α e crer em não-α. Percebe-se, então, que há apenas uma intersecção entre os conjuntos de intenções e de crenças de um dado agente ao invés do segundo conter o primeiro (caso da suposição forte referenciado anteriormente na lógica de Cohen e Levesque e concordada por Rao e Georgeff em [RG91, RG93]).

Note que possíveis relações entre desejos e crenças não são evidenciadas, pois os desejos podem ser conflitantes ou contraditórios. Apesar de ser interessante e necessário estabelecer relações entre desejos e intenções, isto não será feito agora.

Compromisso

Falou-se na sub-seção anterior sobre compromisso. O compromisso com a intenção é necessária para o agente conseguir realizá-la. Pode-se pensar num exemplo em que a intenção seja seguir carreira acadêmica e, para isso, deve-se fazer a graduação, pós e o concurso para professor. Caso não haja comprometimento do agente para com a intenção, ele desistirá da mesma antes mesmo de se tornar um formando.

Há na literatura 3 estratégias de compromisso. São elas: Compromisso cego (*"Blind Commitment"*) em que o agente mantém a intenção até acreditar que esta foi realizada; "Compromisso simples" (*"Single-minded Commitment"*) em que o agente mantém a intenção até acreditar que foi realizada ou que é impossível realizá-la; e "Compromisso de mente-aberta" (*"Open-minded Commitment"*) em que o agente mantém a intenção até ele acreditar que a satisfez ou enquanto ele deseje satisfazê-la.

2.3.3 Implementando um Agente BDI

Para um melhor entendimento, estudar-se-á como seria implementado um agente BDI de uma forma geral, isto é, serão apresentados os passos que não podem deixar de existir. Na modelagem de um agente necessariamente precisa: observar o mundo para posteriormente construir seu conjunto de crenças e a partir de então deliberar suas intenções e determinar o seu plano de ações. Assim, um algoritmo simplificado de um agente BDI poderia ser da seguinte forma[8]:

Observando a figura 2.1 atenta-se aos passos 4 e 5, já que o tempo de deliberação ($t_{delibera}$) e o tempo do raciocínio meio-fim ($t_{meio-fim}$) não são tempos instantâneos e com isso vem a preocupação da possibilidade de ocorrer alguma mudança no mundo no meio do intervalo de tempo de $t_{delibera}$ ou de $t_{meio-fim}$. Este problema não é tratado no algoritmo acima, mas pode ser solucionado a partir de reconsiderações das intenções quando necessário.

[7] Essa tese será discutida com mais detalhes no sexto capítulo

[8] Os algoritmos apresentados nesta subseção são extraídos de [Woo02a, Woo00a].

```
while true {
    observa o mundo
    atualiza o modelo de mundo interno
    deliberação das intenções a serem realizadas
    usa o raciocínio meio-fim para estabelecer o plano
    executa o plano
    }
```

Figura 2.1: Controle de "loop" básico de um agente BDI.

Função	**Assinatura**
frc:	$\wp(Bel) \times Per \rightarrow \wp(Bel)$
delibera:	$\wp(Bel) \rightarrow \wp(Int)$
plano	$\wp(Bel) \times \wp(Int) \rightarrow Plano$

Tabela 2.5: Formalização das funções de definição de crenças intenções e planos.

Vê-se também no algoritmo acima o "plano". Neste trabalho, será adotada a conceitualização proposta em [Woo00a]: planos são "receitas" para a realização de intenções. Eles são constituídos por pré-condições (condições necessárias para o plano ser executado), pós-condições (estados os quais a realização do plano irá levar o agente) e o corpo do plano, onde está alocado a receita atual (uma sequência de ações). Um exemplo para se ter uma boa visualização da estrutura de um plano seria o seguinte: supondo que o plano do agente é pegar um taxi para ir ao aeroporto então a pré-condição seria ele ter fundos para pagar o taxi, o corpo seria a sequência de ações executadas pelo agente e a pós-condição é ele estar no aeroporto (idealmente) e ter menos dinheiro.

Remodelando o algoritmo acima com as funções descritas na tabela 2.5[9], obtem-se o algoritmo da figura 2.2.

```
B := B0
while true {
    obtém próxima percepção ρ
    B := frc(B, ρ)
    I := delibera(B)
    π := plano(B, I)
    executa π
    }
```

Figura 2.2: Segunda versão do algoritmo que implementa um agente BDI.

[9]"frc" é uma função de revisão de crença e $\wp(Bel)$, Per e $\wp(Int)$ são respectivamente os conjuntos das partes das crenças, percepções e intenções do(s) agente(s).

Função	**Assinatura**
opções:	$\wp(Bel) \times \wp(Int) \rightarrow \wp(Des)$
filtro	$\wp(Bel) \times \wp(Des) \times \wp(Int) \rightarrow \wp(Int)$

Tabela 2.6: Formalização das funções de definição de desejos e redefinição das intenções.

```
B:=B0
I := I0
  while true {
    obtém próxima percepção ρ
    B := frc(B, ρ)
    D := opções(B, I)
    I := filtro(B, D, I)
    π := plano(B, I)
   executa π
  }
```

Figura 2.3: Terceira versão do algoritmo que implementa um agente BDI.

Refinando um pouco mais o algoritmo da figura, redefine-se a função delibera em duas novas funções (tabela 2.6) e obtem-se o algoritmo da figura 2.3.

A função *opções* é responsável por gerar um conjunto de possíveis alternativas para solução do problema, enquanto a função *filtro* escolhe uma dentre as alternativas, determinadas por *opções*, para se comprometer.

Nessas três versões de algoritmos que implementam a arquitetura interna de agente BDI, nada foi implementado com relação ao tipo de estratégia ou reconsideração de intenções, e verificação de mudança de mundo. No algoritmo da figura 2.4 – implementado com a ajuda da função booleana *reconsidera(**I**,**B**)* a qual analisa as crenças e intenções e retorna verdadeiro se o agente deve reconsiderar as intenções e falso caso contrário. A função booleana *correto(π,**I**,**B**)* a qual determina se o plano está correto ou não com relação às crenças e às intenções – ver-se-á a implementação do "compromisso simples" e a reavaliação de mudanças de intenções depois de cada ação realizada (visto que $\alpha :=$ início(π); e início(π) é apenas a primeira ação de π).

Observando o algoritmo da figura 2.4, discute-se qual seria a melhor estratégia para fazer a verificação da reconsideração de intenções. Com qual frequência deve ser feita tal verificação e por que não utilizar a própria função filtro para determinar, se necessário, um novo conjunto de intenções?

Ao investigar o problema, Kinny e Georgeff em [KG91] analisaram duas estratégias e verificaram a eficiência de: reconsiderar intenções depois de cada plano; e reconsiderar intenções depois de cada ação. Observaram que quanto mais dinâmicos são os ambientes, mais eficiente se torna a segunda opção e a medida que o ambiente se torna mais estático, a reconsideração das intenções depois de cada plano torna o algoritmo mais eficiente.

Para a segunda pergunta, a resposta é bem simples. Wooldridge (em [Woo00a]) observa que o "filtro" não deve ser usado como oráculo para saber se as intenções mudaram ou não, pois tal função é mais custosa do que a função "reconsidera".

```
B:=B0
I := I0
while true {
      obtém próximo ρ
      B := frc(B, ρ)
      D := opções(B, I)
      I := filtro(B, D, I)
      π := plano(B, I)
      while not (vazio(π) or bem-sucedido(I,B) or impossível(I,B)) {
         α := início(π)
         executa(α)
         π := resto(π)
         obtém próxima percepção ρ
         B := frc(B, ρ)
         if reconsidera(I, B) then {
            D := opções(B, I)
            I := filtro(B, D, I)
         }
         if not (correto(π, I, B)) then
            π := plano(B, I)
      }
}
```

Figura 2.4: Quarta versão da implementação de um agente BDI.

Note que na aplicação BDI, o desejo é na verdade uma pró-atitude menor (estados de mundo nos quais o agente quer chegar para conseguir alcançar sua intenção) originada das próprias intenções e suas crenças. O que é bem natural de se pensar uma vez que estes estados são parte da receita a qual deve ser seguida para alcançar a intenção. Diferentemente do proposto por [RG91, RG93] quando definem que as intenções são os desejos, os quais, o agente se compromete em realizar sugerindo assim que os desejos antecedem as intenções.

2.4 Considerações Finais

Durante o capítulo viu-se três formas de se formalizar um agente ou um sistema multiagente cognitivo. Cabe agora argumentar o porquê da importância de formalizar logicamente uma teoria de ações para agentes inteligentes. Cohen e Levesque, em [CL90], fazem algumas considerações quanto a isso. Primeiramente, eles argumentam que especificar formalmente o *design* de agentes autônomos contribui com a pesquisa em inteligência artificial (IA). As análises com relação à teoria de ações acabam por capturar os conceitos ordinários de intenção. E concentrar a pesquisa de IA em algoritmos cuja função é achar planos para realizar determinados objetivos ou a monitoração da execução de planos e replanejamento, não é suficiente. Frenquentemente, tais pesquisas têm ignorado assuntos como o balanço racional, por exemplo, além de que a teoria de ações intencionais são expressadas apenas por código (ou no máximo por uma especificação de software) enquanto as relações entre crenças, desejos, planos e ações ficam implícitas no próprio código. Quanto à importância das formalizações para a filosofia, conforme Cohen e Levesque afirmam, as formalizações são necessárias para esclarecer os conceitos envolvidos na teoria e para melhor entender a relação entre tais conceitos. Linder, Hoek e Meyer ([vLvdHM98]) afirmam ainda que para tal assunto (comportamento ou cognição dos agentes) há uma maior necessidade de formalização, pois lida com o senso-comum de conceitos como conhecimento, crença, habilidade, etc. gerando assim várias formas de conceitualização e interpretação o que consequentemente provoca ambiguidades as quais devem ser eliminadas.

Todas as formalizações, apesar de haver grandes diferenças entre elas, utilizam a lógica modal (ou simplesmente modalidades como no caso da formalização de Cohen e Levesque – [CL90]). A lógica modal pode ser uma lógica intencional[10], o que permite criar uma formalização aceitável das noções intencionais com um esforço bem menor do que, por exemplo, se fosse feito pela lógica clássica de predicados[11].

O modelo BDI se torna mais interessante do que os outros por ele estar fundamentado sobre um componente filósofico reconhecidamente aceitável e da comprovação de que o mesmo é computável em função dele já ter sido implementado inúmeras vezes.

É importante falar ainda da lógica dos jogos a qual faz uso de técnicas da teoria dos jogos ([OR94] – muito aplicado também na área de teorias econômicas). Vários trabalhos mencionam ou trabalham a ligação entre as estruturas de Kripke e os jogos, como exemplo: [Par85], que trabalha sobre o escopo da lógica dinâmica dos jogos; [Non01], o qual estuda as relações entre as lógicas temporais com tempos ramificados e modelos de jogos; [HvdHMW02], que analisa através da lógica modal as noções de teoria dos jogos, do equilíbrio de Nash e a melhor estratégia de resposta de agentes e, por fim, cita-se a lógica de Coalizão de Pauly ([Pau02]) que é uma lógica modal na qual há um conjunto de operadores na forma $[C]$. Uma fórmula $[C]\varphi$ significa que o conjunto de agentes C assegura que φ é verdadeiro.

Uma vantagem da teoria dos jogos é sua reconhecida aplicação na computação [Bin92], entretanto, assim como a teoria de decisão que utiliza probabilidade para determinar qual será o comportamento desejado para o agente e possui complexidade exponencial com relação as possíveis saídas (resultados das performances do agente), a teoria dos jogos tem limitações quantitativas e também, segundo [Woo00a], com respeito a sua "capacidade" de determinar o que fazer sem especificar com clareza como fazer.

Nos próximos dois capítulos estudar-se-ão as lógicas modal e temporal (capítulo 3) e a teoria fuzzy (capítulo 4) as quais são fudamentais para o entendimento da lógica BDI a ser apresentada e da proposta de lógica BDI fuzzy.

[10] O termo "intencional" referido à lógica tem o mesmo significado de "intencionalidade" (no sentido comum e não filosófico) definido em [Den97]. Ou seja, lógica intencional é uma lógica na qual são inferidas propriedades envolvendo a atitude mental intenção.

[11] Em [Woo00a, 13] apresenta-se um bom exemplo da dificuldade de representar uma situação prática que envolve ações, utilizando a lógica de primeira ordem e a facilidade de fazê-la usando a ferramenta modal.

Capítulo 3

Lógicas Modais e Temporais

Será introduzido neste capítulo os temas lógica modal de predicado e lógica temporal, resumindo assim suas principais características. Este capítulo trata apenas de um texto introdutório sobre o assunto pois não há necessidade de se estender no mesmo, neste trabalho.

3.1 Introdução

A lógica das sentenças modais foi inicialmente discutida por Aristóteles e por lógicos medievais. Também obteve avanços lógicos e filosóficos através das contribuições de MacColl em 1880 e 1906. Mas, foi o desenvolvimento formal do cálculo clássico (não-modal) por Frege e Russel que motivou Lewis, em 1918, às primeiras axiomatizações da lógica modal, e Marcus, em 1946, estendeu à lógica modal de predicados.

Segundo William e Marta Kneale ([KK62]), Frege em *Begriffsschrift* (livro no qual Frege formaliza a lógica) eliminou as distinções modais, pois achava as modalidades irrelevantes para o seu fim. Para Frege, expressões como "ter que", "poder", "necessário" e "possível" envolvem uma referência oblíqua ao conhecimento humano e, por isso, seria um erro estabelecer regras de inferência aplicáveis apenas a operadores modais, ou seja, estes não devem fazer parte de uma lógica pura. No entanto, as noções de necessidade e possibilidade não pertencem à epistemologia ou à outra ciência qualquer como Frege julgava.

E segundo Susan Haack ([Haa02]), Lewis foi motivado pela sua insatisfação com a definição de "implicação material" – ponto central da lógica do *Begriffsschrift* ([Fre79]) e do *Principia Mathematica* de Russel ([Rus03]) – para então criar os operadores $\Diamond$ (possível) e $\Box$ (necessário) os quais deram início à formalização da lógica modal. Lewis afirmava que a implicação material da lógica clássica é inadequada uma vez que 'p' implica materialmente 'q' se 'p' é falso ou 'q' é verdadeiro; e para Lewis, a noção intuitiva da implicação requer que 'p' não possa ser verdadeiro e 'q' falso. Lewis então criou três novos operadores: a "implicação estrita", o "possível" e o "necessário" definidos respectivamente da seguinte forma[1]:

$$A \Longrightarrow B \stackrel{\text{def}}{=} \neg\Diamond(A \wedge \neg B) \equiv \Box(A \rightarrow B) \tag{3.1}$$

[1] Observe que ou '$\Diamond$' ou '$\Box$' é primitivo. De forma que se '$\Box$' for primitivo, '$\Diamond$' é definido através da definição 3.2, caso contrário, '$\Box$' é definido através da definição 3.3.

$$\Diamond A \stackrel{\text{def}}{=} \neg\Box\neg A \tag{3.2}$$

$$\Box A \stackrel{\text{def}}{=} \neg\Diamond\neg A \tag{3.3}$$

Hoje, diversos são os campos que se utilizam da lógica modal. A lógica modal é uma das principais ferramentas lógicas para áreas da computação como sistemas multiagentes e especificações e verificações de programas.

A lógica modal é vista atualmente como uma extensão da lógica clássica no sentido de que a primeira acrescenta à linguagem do último dois novos operadores ($\Box$ e $\Diamond$), bem como, novos teoremas e regras de inferência. Por isso, antes de conhecer a teoria formal modal, ver-ser-á a teoria formal clássica de predicados para posteriormente estendê-la à uma teoria formal modal de predicados.

3.2 Lógica Clássica de Predicados de 1ª Ordem

A lógica clássica de predicados de 1ª ordem surgiu da necessidade de uma maior expressibilidade a qual a lógica proposicional clássica não era capaz de oferecer. Pois, esta não é suficientemente expressiva para representar perfeitamente em sua linguagem formal expressões do tipo "todos os seres vivos vão morrer" ou "não existem jacas lisas". Para representar tais sentenças é desejável fazer-se uso de predicados, tais como: LISAS(x) e VAI_MORRER(x), e dos quantificadores para todo ($\forall$) e existe ($\exists$). Assim as sentenças seriam escritas em linguagem formal de 1ª ordem, respectivamente, por

$$\forall x.(\text{SER\_VIVO}(x) \rightarrow \text{VAI\_MORRER}(x)$$
$$\neg\exists x.(\text{JACAS}(x) \wedge \text{LISAS}(x)).$$

Há vários detalhes acerca da construção de sentenças em linguagem de predicados e seus diferentes significados, bem como, a interpretação do $\forall$ e o $\exists$ como sendo, respectivamente, uma conjunção e uma disjunção de predicados ambas infinitas, no entanto, tais detalhes não serão tratados aqui. Preocupar-se-á apenas em mostrar a teoria formal dessa lógica[2].

3.2.1 Linguagem de 1ª Ordem

A linguagem de 1ª ordem deve ser capaz de considerar o valor de verdade de sentenças construídas a partir de sentenças atômicas as quais ora são falsas ora são verdadeiras dependendo do valor do universo de domínio, por exemplo: "x é brasileiro" é uma asserção atômica que poder ser verdadeira ou falsa, pois ela é representada

[2] A linguagem de 1ª ordem, os axiomas e regras foram definidos segundo o estilo de Benjamín e Acióly Bedregal em [BA07]

em linguagem de 1ª ordem a partir de um predicado aplicado a um argumento o qual pode assumir diferentes valores.

Como nas linguagens de 1ª ordem o alfabeto (Σ) contém variáveis, símbolos de função, símbolos de predicados e operadores lógicos, então, ele deve ser definido a partir de Σ-estruturas. A linguagem de 1ª ordem, a ser usada neste trabalho, tem seu alfabeto definido a seguir:

$$\Sigma = X \cup \Sigma_C \cup \Sigma_F \cup \Sigma_R \cup \Sigma_L \cup \Sigma_P$$

onde

1. $X = \{x, y, z, x_1, x_2, \ldots, y_1, y_2, \ldots, z_1, z_2, \ldots\}$ é um conjunto enumerável de símbolos de variáveis
2. $\Sigma_C = \{a, b, c, a_1, a_2, \ldots, b_1, b_2, \ldots, c_1, c_2, \ldots\}$ é um conjunto enumerável de símbolos de constantes
3. $\Sigma_F = \{f, f_1, f_2, f_3 \ldots\}$ é um conjunto enumerável de símbolos de funções
4. $\Sigma_R = \{P, P_1, P_2, P_3 \ldots\}$ é um conjunto enumerável de símbolos de relações ou predicados
5. $\Sigma_L = \{\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \forall, \exists\}$ é o conjunto de símbolos lógicos
6. $\Sigma_P = \{(,), ., ,\}$ é um conjunto de símbolos de pontuação.

Σ_L é o mesmo para qualquer linguagem de predicados de 1ª ordem (a não ser que seja feita a simplificação desta estrutura e se defina uns operadores derivados de outros), variando o que será chamado de Σ_d, onde $\Sigma_d = \Sigma_F \cup \Sigma_R$. Seja $aridade : \Sigma_d \longrightarrow \mathbb{N}$ uma função que indica a aridade dos símbolos de função e de predicado.

Para a definição da linguagem formal de predicados de 1ª ordem ($Ling(\mathcal{L}^P)$) primeiramente se constrói a linguagem dos termos de predicados ($Ling(\mathcal{L}^T)$) a qual é usada em $Ling(\mathcal{L}^P)$ como parte do seu alfabeto.

Definição 3.2.1 *A linguagem dos termos de 1ª ordem é a linguagem formal $\mathcal{L}^T$ tal que $\mathcal{L}^T = \langle \Sigma, \mathcal{G}^T \rangle$ onde $\Sigma = X \cup \Sigma_C \cup \Sigma_F \cup \{(,),,\}$ e $\mathcal{G}^T = \{T_1, T_2, T_3\}$:*

T_1 $\dfrac{-}{x}$, $x \in X$

T_2 $\dfrac{-}{a}$, $a \in \Sigma_C$

T_3 $\dfrac{\tau_1, \ldots, \tau_n}{f(\tau_1, \ldots, \tau_n)}$, $f \in \Sigma_F$ *e* $aridade(f) = n$.

E a linguagem de 1ª ordem é a linguagem formal $\mathcal{L}^P$ tal que $\mathcal{L}^P = \langle \Sigma, \mathcal{G} \rangle$ onde $\Sigma = Ling(\mathcal{L}^T) \cup \Sigma_R \cup \Sigma_L \cup \Sigma_P$ e $\mathcal{G}^P = \{F_1, \ldots, F_8\}$. Sendo as regras $F_1, \ldots, F_8$ mostradas abaixo.

F_1 $\dfrac{\tau_1, \ldots, \tau_n}{R(\tau_1, \ldots, \tau_n)}$, $R \in \Sigma_R$, $aridade(R) = n$ *e* $\tau_j \in Ling\mathcal{L}^T)$, $\forall j.(1 \leq j \leq n)$

F_2 $\dfrac{\varphi}{\neg\varphi}$

F_3 $\dfrac{\varphi,\ \phi}{(\varphi \vee \phi)}$

F_4 $\dfrac{\varphi,\ \phi}{(\varphi \wedge \phi)}$

F_5 $\dfrac{\varphi,\ \phi}{(\varphi \rightarrow \phi)}$

F_6 $\dfrac{\varphi,\ \phi}{(\varphi \leftrightarrow \phi)}$

F_7 $\dfrac{\varphi}{(\exists x.\varphi)},\ x \in X$

F_8 $\dfrac{\varphi}{(\forall x.\varphi)},\ x \in X$

Observação 3.2.1 *1. O conjunto L^P é o conjunto de fórmulas bem-formadas (fbf) de $\mathcal{L}^P$. Ou seja, $L^P = Ling(\mathcal{L}^P)$. As fbf de qualquer que seja a linguagem apresentada neste trabalho, serão denotadas por $\{\varphi,\ \phi,\ \psi,\ \varphi_1,\ \varphi_2,\ \varphi_3 \ldots,\ \phi_1,$ $\phi_2,\ \phi_3, \ldots, \psi_1,\ \psi_2,\ \psi_3, \ldots\}$.*

2. Nas regras gramaticais φ e ϕ são meta-variáveis e por isso podem ser substituídas por qualquer fórmula de $\mathcal{L}^P$.

3. As fórmulas obtidas pela regra F_1 são as chamadas fórmulas atômicas e o conjuntos de fórmulas atômicas de L^P será denotado por L_0^P.

4. Os conectivos são ordenados da mais baixa precedência para a mais alta como segue: $\leftrightarrow$, $\rightarrow$, $\wedge$, $\vee$, $\neg$. E os parênteses podem ser eliminados seguindo a regra de que $\neg$ tem a mais alta precedência e $\leftrightarrow$ a mais baixa.

3.2.2 Semântica da Lógica de 1ª Ordem

Definida a linguagem, é necessário definir interpretações para as fórmulas de $\mathcal{L}^P$. Seja V uma função que mapeia uma fbf de L^P nos valores verdadeiro (denotado por 1) e falso (denotado por 0). Logo $V : L^P \rightarrow \mathbb{B}$, onde $\mathbb{B} = \{0, 1\}$. Daí, segue a definição:

Definição 3.2.2 *Uma interpretação $\mathcal{I}$ para o domínio, ou universo de discurso, I é uma função de tal forma que*

$$\begin{array}{l} \forall f \in \Sigma_F,\ I(f) : D^{aridade(f)} \rightarrow D \\ \forall P \in \Sigma_R,\ I(P) \subseteq D^{aridade(P)} \\ \forall a \in \Sigma_C,\ I(a) \in D \end{array} .$$

Daí, para determinar o valor verdade de qualquer fbf de L^P dada uma interpretação $\mathcal{I}$ e atribuições ρ de valores de I às variáveis ($\rho : X \rightarrow I$); primeiro determina-se os valores dos termos e, em seguida, interpreta-se as fórmulas. Em outra palavras, cada ρ determina uma função de valoração $\mathcal{I}_\rho : Ling(\mathcal{L}^T) \to I$ e uma $V_\rho : Ling(\mathcal{L}^P) \to \{0,1\}$ as quais são definidas recursivamente como segue:

$$\mathcal{I}_\rho(x) = \rho(x) \text{ para cada } x \in X$$
$$\mathcal{I}_\rho(a) = \mathcal{I}(a) \text{ para cada } a \in \Sigma_C$$
$$\mathcal{I}_\rho(f(\tau_1, \ldots \tau_n)) = I(f)(\mathcal{I}_\rho(\tau_1), \ldots, \mathcal{I}_\rho(\tau_n)))$$
$$\mathcal{I}_\rho(P(\tau_1, \ldots, \tau_n)) = I(P)(\mathcal{I}_\rho(\tau_1), \ldots, \mathcal{I}_\rho(\tau_n)) \text{ para cada } P \in \Sigma_R \text{ com } aridade(P) = n$$
$$\text{e } \tau_1, \ldots, \tau_n \in Ling(\mathcal{L}^T)$$

$$V_\rho(\neg\varphi) = \begin{cases} 1 & \text{, se } V_\rho(\varphi) = 0 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\varphi \vee \phi) = \begin{cases} 1 & \text{, se } V_\rho(\varphi) = 1 \text{ ou } V_\rho(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\varphi \wedge \phi) = \begin{cases} 1 & \text{, se } V_\rho(\varphi) = 1 \text{ e } V_\rho(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\varphi \to \phi) = \begin{cases} 0 & \text{, se } V_\rho(\varphi) = 1 \text{ e } V_\rho(\phi) = 0 \\ 1 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\varphi \leftrightarrow \phi) = \begin{cases} 1 & \text{, se } V_\rho(\varphi) = V_\rho(\phi) \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\forall x.\phi) = \begin{cases} 1 & \text{, se para toda atribuição } \rho' : X \to I \text{ tal que } \rho'(y) = \rho(y) \\ & \forall y.(y \neq x),\ V_{\rho'}(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_\rho(\exists x.\phi) = \begin{cases} 1 & \text{, se para alguma atribuição } \rho' : X \to I \text{ tal que } \rho'(y) = \rho(y) \\ & \forall y.(y \neq x),\ V_{\rho'}(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

Definição 3.2.3 *Seja $\mathcal{I}$ uma interpretação e ρ uma atribuição para as variáveis. Uma fórmula φ é verdadeira para uma ρ em um I, denotado por $\mathcal{I} \models_\rho \varphi$, se $V_\rho(\varphi) = 1$. Neste caso $(\mathcal{I}, \rho)$ é dito um modelo de φ – uma fórmula é dita satisfazível se existe algum modelo. E φ é falsa para uma ρ em uma $\mathcal{I}$, denotado por $\mathcal{I} \not\models_\rho \varphi$, se $V_\rho(\varphi) = 0$. Uma fórmula pode ainda ser verdadeira em uma $\mathcal{I}$, denotado por $\mathcal{I} \models \varphi$, se para qualquer ρ temos que $\mathcal{I} \models_\rho \varphi$ – ou uma contradição, se ela é falsa em uma $\mathcal{I}$, denotado por $\mathcal{I} \not\models \varphi$, se para qualquer ρ $\mathcal{I} \not\models_r ho\varphi$; ou ainda pode ser dita contingente numa $\mathcal{I}$, se existem atribuições ρ e ρ' tal que $I \models_\rho \varphi$ e $I \not\models_{\rho'} \varphi$. Por fim, φ é dita uma fórmula universalmente válida quando ela é verdadeira para qualquer $\mathcal{I}$ e para qualquer ρ, denotado por $\models \varphi$ – o conjunto das fórmulas universalemente válidas de $\mathcal{L}^P$ é definido por $FUV = \{\varphi \in L^P / \models \varphi\}$ –; e é dita insatisfazível, quando não existe um modelo para a mesma, isto é, para qualquer $\mathcal{I}$ e ρ, φ é falsa, denotado por $\not\models \varphi$.*

Da mesma forma que há a definição de modelo de uma fórmula, define-se o modelo de um conjunto de fórmulas Γ de L^P como sendo uma interpretação $\mathcal{I}$ e uma atribuição ρ a qual todas as fórmulas $\varphi \in \Gamma$ são verdadeiras, ou seja, $V_\rho(\varphi) = 1$. Dada as definições anteriores, pode-se definir o conceito de consequência semântica e lógica de 1ª ordem:

Definição 3.2.4 *Seja a linguagem de 1ª ordem $\mathcal{L}^P$ e $\varphi \in L^P$. φ é uma consequência semântica de Γ, denotado por $\Gamma \models \varphi$ se todo modelo de Γ é um modelo de φ. Então, a lógica de 1ª ordem, denotada por Log^P é definida através de uma estrutura munida de sua linguagem de 1ª ordem e a consequência semântica definida anteriormente:*

$$Log^P = \langle \mathcal{L}^P, \models \rangle$$

3.2.3 Teoria Formal da Lógica Clássica de Predicados de 1ª Ordem

A teoria formal de predicados apresentada aqui é constituída por uma linguagem de 1ª ordem simplificada $\mathcal{L}^P_{\rightarrow,\forall,\neg} = \langle \Sigma - \{\wedge, \vee, \leftrightarrow, \exists\}, \mathcal{G}^P_{\rightarrow,\forall} \rangle$, onde $\mathcal{G}^P_{\rightarrow,\forall} = \{F_1, F_2, F_5, F_8\}$; um conjunto de **Axiomas** (Δ):

A_1 $\varphi \rightarrow (\phi \rightarrow \varphi)$

A_2 $(\varphi \rightarrow (\phi \rightarrow \psi)) \rightarrow ((\varphi \rightarrow \phi) \rightarrow (\varphi \rightarrow \psi))$

A_3 $(\neg\varphi \rightarrow \neg\phi) \rightarrow ((\neg\varphi \rightarrow \phi) \rightarrow \varphi)$

A_4 $\forall x.(\varphi \rightarrow \phi) \rightarrow (\varphi \rightarrow \forall x.\phi)$, onde x não ocorre livre em φ

A_5 $\forall x.\varphi \rightarrow \varphi[t/x]$, onde t é livre para x em φ e $\varphi[t/x]$ significa a substituição de todas as ocorrências livres de x pelo termo t em φ;

e **Regras de Inferência:**

Modus Ponens (MP):

$$\frac{\varphi,\ \varphi \rightarrow \phi}{\phi}$$

Generalização Universal (Gen):

$$\frac{\varphi}{\forall x.\varphi}$$

Observação 3.2.2 *1. Dada uma fórmula $\varphi \in Ling(\mathcal{L}^P_{\rightarrow,\forall,\neg})$, diz-se uma consequência sintática na teoria formal $T^P = \langle \mathcal{L}^P_{\rightarrow,\forall,\neg}\{A_1, \ldots, A_5\}, \{MP, Gen\} \rangle$, de um conjunto $\Gamma \subseteq L^P$, denotado por $\Gamma \vdash \varphi$, se existir uma prova de φ a partir de Γ. Ou seja, uma sequência $\varphi_1, \ldots, \varphi_n \in \mathcal{L}$ tal que $\varphi_n = \varphi$ e para cada φ_i, $1 \leq i \leq n$ ou é um axioma ou é uma premissa (no caso, alguma fórmula de Γ) ou é uma consequência direta de φ_j anteriores usando uma das regras de inferência*[3].

[3]Tais definições de consequência sintática e de prova se repete para todas as teorias formais, mudando apenas a linguagem, o conjunto de axiomas e de regras, isto é, mudando apenas a teoria formal sobre a qual a consequência sintática e a prova estão sendo definidas.

2. *Um teorema de uma teoria formal* $T = \langle \mathcal{L}, Ax, \Re \rangle$ *(onde* Ax *é um conjunto de axiomas e* $\Re$ *é um conjunto de regras de inferência) é um elemento* $\varphi \in Ling(\mathcal{L})$ *tal que existe uma prova de* φ *em* T.

3. *A lógica apresentada pela teoria formal de 1ª ordem* T^P*, denotado por* $\mathcal{T}(T^P)$ *é definida como sendo o conjunto de todos os teoremas da teoria formal* T^P*:*

$$\mathcal{T}(T^P) = \{\varphi \in L^P_{\rightarrow,\forall} \ / \ \vdash \varphi\}.$$

4. *Teoremas envolvendo quantificador existencial, conjunção, disjunção e bi-implicação podem ser provados usando as respectivas equivalências:*

   4.a $\exists x.\varphi \equiv \neg\forall x.\neg\varphi$

   4.b $\varphi \wedge \phi \equiv \neg(\varphi \rightarrow \neg\phi)$

   4.c $\varphi \vee \phi \equiv \neg\varphi \rightarrow \phi$

   4.d $\varphi \leftrightarrow \phi \equiv (\varphi \rightarrow \phi) \wedge (\phi \rightarrow \varphi)$

3.3 Linguagem da Lógica Modal de 1ª Ordem

Como havia sido dito, a extensão da lógica de 1ª ordem para a lógica modal de 1ª ordem com respeito às suas linguagens foi o acréscimo de dois novos operadores à linguagem da lógica de 1ª ordem. Então, a linguagem da lógica modal de 1ª ordem, denotada por $\mathcal{L}^M$, onde $\mathcal{L}^M = \langle \Sigma^M, \mathcal{G}^M \rangle$, sendo $\Sigma^M = X \cup \Sigma_d \cup \Sigma_{L'} \cup \Sigma_P$ onde $\Sigma_{L'} = \{\neg, \vee, \wedge, \rightarrow, \leftrightarrow, \forall, \exists, \Box, \Diamond\}$ – e $\mathcal{G}^M = \{F_1, \ldots, F_{10}\}$, sendo F_9 e F_{10} definidos a seguir:

F_9 $\dfrac{\varphi}{\Box\varphi}$

F_{10} $\dfrac{\varphi}{\Diamond\varphi}$.

Observação 3.3.1 *L^M é o conjunto de fbf de $\mathcal{L}^M$. E L^M_0 é o conjunto de fórmulas atômicas de $\mathcal{L}^M$.*

3.4 Semântica dos Mundos Possíveis

A semântica, em função da adição das modalidades, também mudou bastante. Na lógica modal é utilizada a semântica dos mundos possíveis. O conceito de mundos possíveis é usado para expressar as características das modalidades. A estrutura dos mundos possíveis é composta por inúmeros mundos (podendo ser composto por um número infinito de mundos). Em cada mundo, há um conjunto de proposições (ou propriedades) as quais são consideradas verdadeiras naquele mundo. Dessa forma, um mundo pode ser visto como um conjunto de fbf as quais são verdadeiras no mesmo. Também é possível que uma certa proposição seja sempre verdadeira a partir de um dado mundo w_1 (verdadeira em todos os mundos relacionados com w_1, essas são chamadas de proposições necessárias) ou que ela seja verdadeira em pelo menos um mundo possível relacionado ao w_1 — chamadas de proposições possíveis.

3.4.1 Modelos, Verdade e Validade na Lógica Modal

As fbf de L^M são operadas sobre uma estrutura (chamada de "Interpretação" em [Mor00] e de *Frame* em [Gar00, Gol92]) composta pelo par (W, R), no qual W é um conjunto de mundos ($W = \{w_1, w_2, w_3, \ldots\}$) e R é uma relação de acessibilidade entre os mundos ($R \subseteq W \times W$). Entretanto, a interpretação dessas fórmulas é feita a partir de um modelo ou semântica de Kripke.

Um modelo ou uma semântica de Kripke (semântica de mundos possíveis) é uma estrutura $M = \langle W, R, V\rangle$, onde a atribuição dos valores verdade das variáveis (livres) são atribuídos em cada mundo w por $\rho_w : X \to D_w$, sendo D_w o domínio (universo de discurso) do mundo w, e que pode ser estendida para $\mathcal{V}_{\rho_w} : Ling(\mathcal{L}^T) \to D_w$[4] da seguinte forma:

$$\mathcal{V}_{\rho_w}(a) = a_w \text{ para cada } a \in \Sigma_C$$
$$\mathcal{V}_{\rho_w}(x) = \rho_w(x) = w(x) \text{ para cada } x \in X$$
$$\mathcal{V}_{\rho_w}(f(\tau_1, \ldots \tau_n)) = f_w(\mathcal{V}_{\rho_w}(\tau_1), \ldots, \mathcal{V}_{\rho_w}(\tau_n))$$

onde a_w é a interpretação do símbolo constante a em w e portanto $a_w \in D_w$. Analogamente, f_w é a interpretação do símbolo de função f em w e portanto $f_w : D_w^{aridade(f)} \to D_w$.

V_ρ é uma função de valoração de verdade que mapeia as sentenças de um determinado mundo em valores do conjunto $\mathbb{B} = \{0, 1\}$ a partir da atribuição ρ, ou seja, $V_\rho : W \times L^M \to \mathbb{B}$. Assim, as condições de verdade das fbf geradas pelas regras $T_1, T_2, T_3, F_1, \ldots, F_8$ são bem semelhantes às mesmas definidas na subseção 3.2.2, no entanto, ao invés de $V_\rho(\varphi)$ estar-se-ia definido $V_{\rho_w}(\varphi)$, onde $V_{\rho_w}(\varphi) = V_\rho(w, \varphi)$ é a denotação para o valor de φ no mundo w.[5]

$$V_w(P(\tau_1, \ldots \tau_n)) = P_w(\mathcal{V}_{\rho_w}(\tau_1), \ldots, \mathcal{V}_{\rho_w}(\tau_n))) \text{ para cada } P \in \Sigma_R \text{ com } aridade(P) = n \text{ e } \tau_1 \ldots \tau_n \in \mathcal{L}^T$$

onde P_w é a interpretação do símbolo de predicados P em w e portanto $P_w \subseteq D_w^{aridade(P)}$.

$$V_w(\neg\varphi) = \begin{cases} 1 & \text{, se } V_w(\varphi) = 0 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\varphi \vee \phi) = \begin{cases} 1 & \text{, se } V_w(\varphi) = 1 \text{ ou } V_w(\phi) = 1 \text{ ou ambos} \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\varphi \wedge \phi) = \begin{cases} 1 & \text{, se } V_w(\varphi) = 1 \text{ e } V_w(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\varphi \to \phi) = \begin{cases} 0 & \text{, se } V_w(\varphi) = 1 \text{ e } V_w(\phi) = 0 \\ 1 & \text{, caso contrário} \end{cases}$$

[4]A linguagem dos termos $Ling(\mathcal{L}^T)$ é definida similarmente como foi definida na seção 3.2.2.

[5]A partir de agora omitir-se-á o ρ quando este for irrelevante ou claro do contexto.

$$V_w(\varphi \leftrightarrow \phi) = \begin{cases} 1 & \text{, se } V_w(\varphi) = V_w(\phi) \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\forall x.\phi) = \begin{cases} 1 & \text{, se para toda atribuição } \rho'_w : X \to D \text{ tal que } \rho'_w(y) = \rho_w(y) \\ & \forall y.(y \neq x)\, V_w(\varphi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\exists x.\phi) = \begin{cases} 1 & \text{, se para alguma atribuição } \rho'_w : X \to D \text{ tal que } \rho'_w(y) = \rho_w(y) \\ & \forall y.(y \neq x)\, V_w(\varphi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_w(\Box\phi) = \begin{cases} 1 & \text{, se para todo } w' \mid wRw',\ V_{w'}(\phi) = 1 \\ 0 & \text{, se para algum } w' \mid wRw',\ V_{w'}(\phi) = 0 \end{cases}$$

$$V_w(\Diamond\phi) = \begin{cases} 1 & \text{, se para algum } w' \mid wRw',\ V_{w'}(\phi) = 1 \\ 0 & \text{, se para todo } w' \mid wRw',\ V_{w'}(\phi) = 0 \end{cases}$$

Definição 3.4.1 *Seja um modelo $M = \langle W, R, V \rangle$, a função ρ e uma fórmula $\varphi \in L^M$. φ é satisfazível para um modelo M (ou M satisfaz φ), se existe um mundo w tal que $V_w(\varphi) = 1$ (seja $w \in W \in M$), denotado por $M, w, V \models \varphi$. E φ é dita insatisfazível para M, denotado por $M, V \not\models \varphi$, se não existe um mundo $w \in W$ do modelo M tal que $V_w(\varphi) = 1$, ou em outras palavras: $V_w(\varphi) = 0$ – $M, w, V \not\models \varphi$ – para todos os $w \in W$. Uma fórmula φ é verdadeira em um modelo, denotado por $M, V \models \varphi$ se para todo $w \in W \in M$, $V_w(\varphi) = 1$, ou seja, φ é satisfeita em todos os mundos w de M; φ é falsa em M sse ela é insatisfazível para M; e φ é contingente se existem mundos w e w' de M tal que $V_w(\varphi) = 1$ e $V_{w'}(\varphi) = 0$, isto é, φ é contigente sse $M, w, V \models \varphi$ e $M, w', V \not\models \varphi$.*

Com isso, pode-se afirmar que uma fbf φ de L^M é universalmente válida sse ela é verdadeira em todos os modelos, denotado por $\models \varphi$. No entanto, a validade também pode ser dada em um dado conjunto de modelos (em uma dada classe), denotado por $C \models \varphi$.

Definição 3.4.2 *Seja a linguagem modal $\mathcal{L}^M = \langle \Sigma^M, \mathcal{G}^M \rangle$ e $\Gamma \subseteq L^M$. φ é uma consequência semântica de Γ, denotado por $\Gamma \models \varphi$ se para todo modelo (um mundo w com suas respectivas relações de acessibilidade e valorações de fórmulas) onde Γ é verdadeiro, φ também o é ($\forall w$ / se $M, w, V \models \Gamma$ então $M, w, V \models \varphi$).*

Finalmente, a lógica modal pode ser definida sobre uma abordagem semântica:

Definição 3.4.3 *Seja a linguagem modal $\mathcal{L}^M = \langle \Sigma^M, \mathcal{G}^M \rangle$ e a consequência semântica $\models$ definida anteriormente. A lógica modal de 1ª ordem definida sobre essas linguagem e inferência é a lógica Log^M onde*

$$Log^M = \langle L^M, \models \rangle.$$

E o conjunto das fórmulas universalmente válidas desta linguagem modal de 1ª ordem é definida como:

$$FMUV = \{\varphi \in L^M \ / \ \models \varphi\}$$

Há meios pelos quais a semântica dos mundos possíveis pode obter certas propriedades. Isto é feito através da adoção de propriedades sintáticas (axiomas), as quais são refletidas semanticamente como propriedades da relação de acessibilidade R. Tais restrições (ou propriedades) serão apresentadas com mais detalhes na seção a seguir.

3.5 Sistemas Formais Modais

Há diferentes tipos de sistemas formais modais, isto é, há inúmeras teorias formais modais. O mais simples deles e que é a base para a construção de todos os outros sistemas formais modais é o sistema K[6].

Observe que todos os teoremas do cálculo de predicados são também logicamente válidos na lógica modal e, portanto, são teoremas da lógica modal. Por isso, os axiomas e regras de inferência do cálculo de predicados devem fazer parte de qualquer que seja o sistema formal modal.

Os sistemas descritos nesta seção também são constituídos pela linguagem $\mathcal{L}^M_{\rightarrow,\neg,\forall,\Box} = \langle \Sigma^M - \{\wedge, \vee, \leftrightarrow, \exists, \Diamond\}, \mathcal{G}^M_{\rightarrow,\forall,\Box}\rangle$ ($\mathcal{G}^M_{\rightarrow,\forall,\Box} = \{F_1, F_2, F_5, F_8, F_{10}\}$), o conjunto de axiomas Δ e as regras de inferência MP e Gen; além da regra Nes:

$$Nes : \frac{\vdash \varphi}{\vdash \Box\varphi} \tag{3.4}$$

Observação 3.5.1 *1. A explicação da regra da Necessitação é que se uma sentença A é universalmente válida, então, ela é verdadeira em todos os mundos possíveis e se ela é verdadeira em todos os mundos possíveis, então, $\Box A$ também é verdadeiro em todos os mundos possíveis, pois seja qual for o mundo possível e quais forem os mundos relacionados com ele, A é verdadeiro.*

2. *A lógica apresentada pela teoria formal modal T^M, denotado por $\mathcal{T}(T^M)$ é definida como sendo o conjunto de todos os teoremas da teoria formal T^M:*

$$\mathcal{T}(T^M) = \{\varphi \in L^M_{\rightarrow,\forall,\Box} / \vdash \varphi\}.$$

3. *Teoremas envolvendo o quantificador existencial, a conjunção, a disjunção, a bi-implicação e a modalidade de possibilidade podem ser provados usando as respectivas equivalências:*

   (a) $\exists x.\varphi \equiv \neg\forall x.\neg\varphi$

   (b) $\varphi \wedge \phi \equiv \neg(\varphi \rightarrow \neg\phi)$

   (c) $\varphi \vee \phi \equiv \neg\varphi \rightarrow \phi$

[6] Recebe o nome de K em homenagem a Saul Kripke

(d) $\varphi \leftrightarrow \phi \equiv (\varphi \rightarrow \phi) \wedge (\phi \rightarrow \varphi)$

(e) $\Diamond\varphi \equiv (\neg\Box\neg\varphi)$

Tanto a lógica modal, quanto a teoria formal modal apresentadas aqui, são as bases para as teorias formais a serem descritas nas sub-seções seguintes. É evidente que cada sistema formal possuirá uma consequência sintática que considerará os axiomas acrescidos em cada teoria, bem como, em uma nova abordagem semântica da lógica, pois ao adicionar um axioma, uma nova condição semântica é considerada (como por exemplo, as relações de acessibilidade serem reflexivas, lineares, etc.).

Os axiomas apresentados a seguir são adicionados aos axiomas da teoria formal clássica de 1ª ordem definida anteriormente neste capítulo.

3.5.1 Sistema Formal K

O sistema K possui além da estrutura descrita anteriormente, mais um axioma: disposto na equação 3.5 (distributividade do $\Box$).

$$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi) \tag{3.5}$$

Mostrar-se-á a seguir os sistemas modais T, D, K4, S4, KB, B e S5. Os sistemas formais aqui descritos, são definidos como o conjunto de teoremas do próprio sistema.

3.5.2 Sistema T

Este sistema possui, além da estrutura descrita anteriormente, os axiomas e regras do sistema K mais o axioma[7] 3.6.

$$\Box\varphi \rightarrow \varphi \tag{3.6}$$

3.5.3 Sistema D

O sistema D possui os axiomas e regras do sistema K mais o axioma 3.7, válido em estruturas com relações de acessibilidade seriais, isto é, estruturas em que sempre há um mundo que se relaciona com o mundo atual.

$$\Box\varphi \rightarrow \Diamond\varphi \tag{3.7}$$

3.5.4 Sistema K4

O sistema K4 possui os axiomas do sistema K mais o axioma 4 (3.8) o qual faz com que as relações de acessibilidade sejam transitivas.

$$\Box\varphi \rightarrow \Box\Box\varphi \tag{3.8}$$

[7] Cada axioma também pode ser chamado de esquema de axioma, uma vez que representa qualquer fórmula com tal forma

3.5.5 Sistema K5

O sistema K5 possui os axiomas do sistema K mais o axioma 5 (3.9) o que torna as relações de acessibilidade euclidianas.

$$\Diamond\varphi \rightarrow \Box\Diamond\varphi \tag{3.9}$$

3.5.6 Sistemas S4, B, KB e S5

O sistema S4 une os axiomas e regras do sistema T mais o axioma 4.

Já o sistema B une os axiomas e regras do sistema T mais o axioma B ou B' (equação 3.10 e equação 3.11, respectivamente). O sistema KB une os axiomas e regras de inferência do sistema K e um dos axiomas B ou B'.

$$\varphi \rightarrow \Box\Diamond\varphi \tag{3.10}$$

$$\Diamond\Box\varphi \rightarrow \varphi \tag{3.11}$$

Por fim, o sistema S5 possui os axiomas e regras do sistema S4 mais o axioma B (ou B'). O S5 pode também ser apresentado através das regras e axiomas do sistema T mais o axioma 5.

3.6 Lógica Temporal

A lógica temporal é um tipo de lógica modal introduzida por Arthur Prior em 1957. Ela é vista como uma extensão da lógica clássica no mesmo sentido de que a lógica modal é uma extensão da lógica clássica. Na lógica temporal, são adicionados ao alfabeto clássico os operadores temporais: '$\Box$' que denota "para qualquer tempo t no futuro"; e '$\Diamond$' o qual denota "para algum tempo t no futuro". Esses operadores são bastante indicados para formalizar sentenças do tipo: "eu estou sempre com fome" e "eventualmente eu estudo". Devido a sua facilidade de expressão de sentenças temporais e por usar a semântica de Kripke, a lógica temporal se torna uma boa ferramenta nas áreas de sistemas dinâmicos, sistemas concorrentes, especificação e verificação de programas, bem como, em sistemas multiagentes e inteligência artificial.

Apesar de se basear na semântica de Kripke, o modelo M é agora definido como uma tupla $\langle T, R, \rho\rangle$, onde T é um conjunto de estados de tempo ($T = \{t0, \ldots tn\}$) em que cada estado t é a representação de um certo tempo, R é o conjunto das relações de acessibilidade entre estados de tempos adjacentes – por isso, a transição de um estado para o outro via relação de acessibilidade ser vista como uma passagem de tempo – e V é uma função que mapeia as fórmulas de L^M presentes em $t \in T$ em $\mathbb{B}$ ($V : T \times L^M \rightarrow \mathbb{B}$).

A linguagem da lógica modal temporal, bem como, a teoria formal modal T^M, definidas anteriormente neste capítulo, são exatamente as mesmas da lógica temporal. No entanto, novas lógicas temporais podem ser criadas através da adição de novos símbolos lógicos ao alfabeto $\Sigma^{L'} \subset \Sigma^M$, tais como:

$\mathcal{W}$ **("a menos que")** $M, t0 \models \varphi \mathcal{W} \phi$ sse $\forall i,\, M, ti \models \varphi$ ou-exclusivo $\exists n \,/\, M, tn \models \phi$ e $\forall i \,/\, 0 \leq i \leq n,\, M, tn \models A$.

$\mathcal{U}$ **("até que")** $M, t0 \models \varphi \mathcal{U} \phi$ sse $\exists n, /\, M, tn \models \phi$ e $\forall i \,/\, 0 \leq i \leq n,\, M, tn \models \varphi$.

$\mathcal{S}$ **("desde")** $M, tk \models \varphi \mathcal{S} \phi$ sse $\exists i \,/ 0 \leq i \leq k,\, M, ti \models \phi$ e $\forall j \,/\, 0 \leq i < j \leq k,\, M, tj \models \varphi$.

$\square_{\triangleleft}$ **("sempre no passado")** $M, t \models \square_{\triangleleft} \varphi$ sse
$\forall t', t'', t''' \ldots /\, t'Rt,\, t''Rt'\, t'''Rt'' \ldots,\, M, (t', t'', t''' \ldots) \models \varphi$

$\lozenge_{\triangleleft}$ **("eventualmente no passado")** $M, t \models \lozenge_{\triangleleft} \varphi$ sse
$\exists t', t'', t''' \ldots /\, t'Rt,\, t''Rt'\, t'''Rt'' \ldots,\, M, (t', t'', t''' \ldots) \models \varphi$

Pode-se ainda definir novos operadores para as lógicas temporais <u>ramificadas</u> (onde $\pi = (t0, \ldots, tn)$ é chamado de caminho):[8]

$\forall\square$, cujo significado é: para todos os caminhos futuros e para todos os estados $t \in \pi$;

$\forall\lozenge$, cujo significado é: para todos os caminhos futuros e em algum estado $t \in \pi$;

$\exists\square$, cujo significado é: para algum caminho π e em todos os estados $t \in \pi$;

$\exists\lozenge$, cujo significado é: para algum caminho π e em algum estado $t \in \pi$.

Com a diversificação dos axiomas podem ainda ser definidas restrições às relações de acessibilidade, podendo gerar diferentes lógicas temporais e modelos de tempo distintos. Ou seja, através das escolhas dos axiomas impõem-se diferentes restrições às relações de acessibilidade, podendo assim definir diferentes modelos de tempo: discreto ou contínuo, linear (todo estado se relaciona com um estado sucessor, além dele mesmo ou não), de futuro ramificado (a partir de um estado de tempo atual há diferentes futuros possíveis), reflexivo (todas as propriedades válidas no futuro são também válidas no presente)...

Na computação, há basicamente a utilização de três tipos de lógicas temporais: LTL (*Linear Tree Logic*); CTL (*Computational Tree Logic*); e CTL*. LTL é uma lógica modal temporal (cujas modalidades se referem ao tempo) proposicional com futuro linear (não ramificado). Dessa forma, o alfabeto de símbolos de uma LTL é composto pelos símbolos lógicos clássicos, mais as modalidades temporais: $\{\bigcirc, \square, \lozenge, \mathcal{U}, \mathcal{W}\}$. Já a CTL é uma lógica temporal de tempo ramificado. Com CTL é possível modelar um futuro não determinado, em outras palavras, há opções de diferentes caminhos a serem tomados, partindo do estado atual. Portanto, além dos operadores modais temporais da LTL, CTL contém ainda os operadores de caminho A e E cujo significado é igual a $\forall\square$ e $\exists\square$ respectivamente. Há outros operadores de caminhos que podem ser adicionados numa CTL. Tais operadores são definidos a partir de um operador de caminhos e uma modalidade temporal (apenas dois operadores), como exemplo, cita-se os operadores $\forall\lozenge$ e $\exists\lozenge$. A CTL* é também uma lógica modal temporal de tempo ramificado semelhante à CTL. No entanto, a CTL* pode construir novos operadores de caminho definidos por mais de dois operadores. Sendo, portanto, a CTL*, uma lógica modal mais expressiva do que a CTL.

[8]Tais definições de símbolos lógicos foram retirados de [Mor00].

As lógicas BDI a serem consideradas nesta dissertação, bem como, a nova lógica BDI desenvolvida neste trabalho, criadas por Wooldridge ([Woo00a]) e Rao e Georgeff ([RG91, RG93]) são formalismos similares à CTL*. Para mais detalhes da CTL*, vide [ES89].

3.7 Considerações Finais

Apresentou-se neste capítulo uma visão introdutória da lógica clássica de predicados, da lógica modal de predicados e da lógica temporal, abordando também os possíveis sistemas modais normais, bem como, alguns modelos de tempo originados do acréscimo (ou restrição) de axiomas. Apesar deste capítulo estudar pouco do tema, pôde-se observar detalhes esclarecedores às definições das formalizações lógicas apresentadas no capítulo anterior. Além de ter construído aqui a base para o desenvolvimento das lógicas BDI e BDI fuzzy.

É importante explicitar também que os sistemas modais podem ser relacionados entre si. O diagrama da figura 3.1 (retirada de [Woo00a]) expressa essas relações, onde um arco de A a B indica que B é um superconjunto estrito de A, isto é, todo teorema de A é um teorema de B e ‘A=B’ significa que A e B contém precisamente os mesmos teoremas.

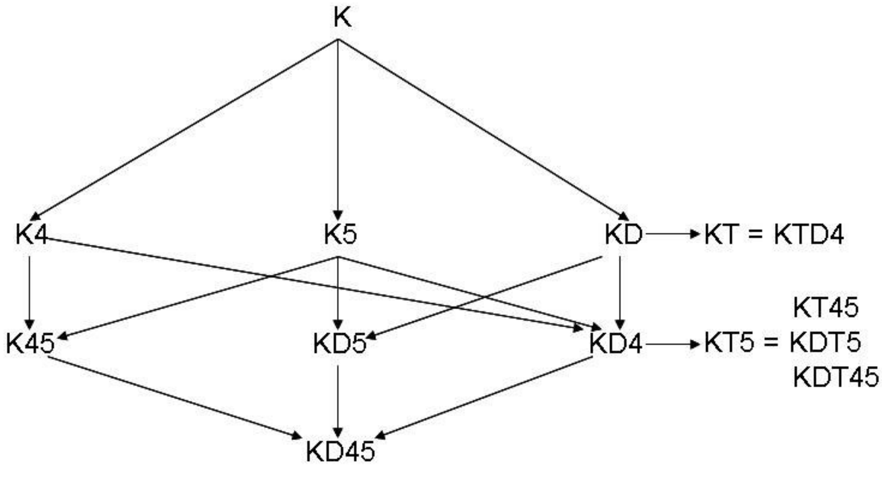


Figura 3.1: Relações entre os sistemas da lógica modal normal.

No capítulo a seguir, ver-se-á a teoria fuzzy (também chamada de nebulosa, ou difusa) de Zadeh para que, em seguida, possa ser estudada a lógica BDI e a proposta de extensão desta.

Capítulo 4

Lógicas BDI

As lógicas BDI são lógicas tri-modais que oferecem as ferramentas necessárias para modelar formalmente um tipo de agente cognitivo. São tri-modais pois seus modelos possuem três relações de acessibilidade a mundos possíveis: relação de "*belief*" (B), "*desire*" (D) e "*intention*" (I).

Fazer-se-á uma explanação sobre as lógicas BDI de Rao e Georgeff ([RG91, RG93]) e a LORA (*Logic of Rational Agents*) de Wooldridge ([Woo00a]). Daí, explicar-se-á o porquê da não utilização de ambas e, posteriormente, (no próximo capítulo) apresentar-se-á uma nova lógica BDI que tem como base essas duas lógicas.

Neste capítulo, apresentar-se-ão duas formalizações lógicas de uma arquitetura de agentes cognitivos seguindo a postura intencional a qual captura muitos pontos da teoria de Bratman e que ao contrário da formalização de Cohen e Levesque, a intenção é uma atitude mental a qual não pode ser definida em função dos desejos e crenças, pois elas são diretamente relevantes à racionalidade do agente como discutido em [Bra87, 19-27] e ela tem a mesma importância de desejos e crenças e, por isso, não pode ser substituída por estes. A formalização de Rao e Georgeff segue uma abordagem formal (teoria formal), mas apresenta também uma semântica informal e condições semânticas correspondente aos axiomas; enquanto LORA segue apenas uma abordagem semântica apresentando quais seriam as propriedades sintáticas correspondentes às inter-relações semânticas entre atitudes mentais (analisando posteriormente qual seria a inter-relação mais aceitável).

Como LORA segue uma abordagem semântica, começar-se-á pela sua apresentação para, em seguida, mostrar a teoria formal BDI de Rao e Georgeff.

4.1 LORA

Wooldridge apresenta sua lógica (LORA) em [Woo00a]. LORA é apresentada didaticamente através de quatro componentes:

- Componente de 1ª ordem (Lógica Clássica de Primeira ordem),

- Componente Temporal (Lógica Temporal),
- Componente BDI,
- Componente de ações.

Há ainda o "componente" de grupos o qual determina quais agentes pertencem a quais grupos, mas como este é bastante simples, ele não será abordado como um componente, mas como um operador: $Agts(i_1 \dots i_n,\, gr)$ o qual significa que os agentes $i_1 \dots i_n$ compõem o grupo gr.

4.1.1 Componente de 1ª Ordem

O primeiro componente é a lógica clássica de predicados de primeira ordem com igualdade. Tal componente é obviamente necessário, pois a lógica modal de 1ª ordem é uma extensão da lógica clássica de 1ª ordem. Seus conectivos devem ser interpretados de acordo com o especificado na tabela 4.1.

As variáveis dentro do escopo dos quantificadores podem pertencer à qualquer um dos domínios **D**$= \langle D_{Ag}, D_{Ac}, D_{Gr}, D_U \rangle$, onde D_{Ag} é o domínio dos agentes, D_{Ac} é o domínio de ações, D_{Gr} é o domínio de grupos de agentes e D_U é um domínio composto por outros indivíduos.

Fórmula	**Interpretação**
$\neg\varphi$	φ não é verdadeiro
$\varphi \wedge \psi$	φ é verdadeiro e ψ é verdadeiro
$\varphi \vee \psi$	φ é verdadeiro ou ψ é verdadeiro
$\varphi \rightarrow \psi$	Se φ é verdadeiro então ψ é verdadeiro
$\varphi \leftrightarrow \psi$	φ é verdadeiro sse ψ é verdadeiro
$\tau = \tau'$	τ é o mesmo que τ'
$\forall x \centerdot \varphi(x)$	Todo objeto x tem a propriedade φ
$\exists x \centerdot \varphi(x)$	Existe um objeto x que tem a propriedade φ

Tabela 4.1: Interpretações dos conectivos da lógica de predicados.

4.1.2 Componente Temporal

Este componente descreve um modelo discreto, onde em cada estado de tempo t há um conjunto de propriedades (sentenças) verdadeiras. Há um estado corrente, o qual o autor o define como "agora" cujo passado é uma sequência linear e discreta de estados. O futuro é ramificado, ou seja, com vários ramos representando os caminhos futuros possíveis ou opcionais (vide a figura 4.1 para entender melhor a estrutura do componente temporal). Uma sequência qualquer de instantes de tempo é chamado de caminho (a figura 4.2 mostra que t0, t1, t2, t5, t9 é um exemplo de caminho); logo, um caminho é uma sequência de estados a qual possui um estado inicial e um estado final (distintos).

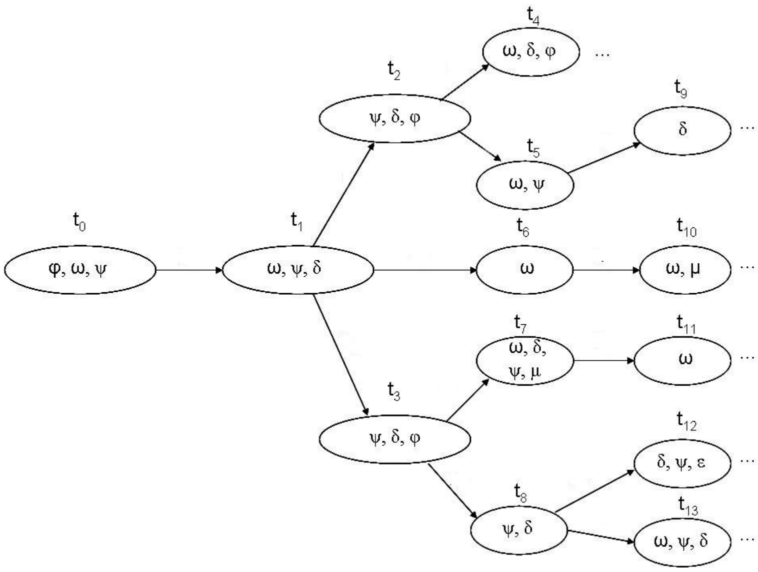


Figura 4.1: Exemplo de uma estrutura temporal ramificada.

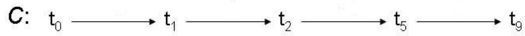


Figura 4.2: Exemplo de um caminho.

Os conectivos temporais e os seus significados são apresentados na tabela 4.2 e 4.3. Em 4.3 são apresentados os quantificadores de caminhos A e E[1] – necessários devido ao futuro ramificado.

Fórmula	**Interpretação**
$\bigcirc\varphi$	φ é verdadeiro apenas no(s) estado(s) seguinte(s) ao atual
$\Diamond\varphi$	φ é verdadeiro em algum estado futuro acessível
$\Box\varphi$	φ é sempre verdadeiro em todos os estados futuros acessíveis
$\varphi\,\mathcal{U}\,\psi$	φ é verdadeiro até que ψ seja verdadeiro
$\varphi\,\mathcal{W}\,\psi$	φ é verdadeiro a menos que ψ seja verdadeiro

Tabela 4.2: Conectivos modais sobre estados.

Fórmula	**Interpretação**
$A(\varphi)$	φ é verdadeiro em todos os caminhos da estrutura temporal
$E(\varphi)$	φ é verdadeiro em algum caminho da estrutura temporal

Tabela 4.3: Quantificadores sobre caminhos.

Observação 4.1.1 *Assume-se que cada um desses estados temporais possuem um passado linear e um futuro ramificado e as relações entre esses estados são também totais e transitivas.*

[1]No terceiro capítulo os operadores A e E foram denotados respectivamente por $\forall\Box$ e $\exists\Diamond$.

Em [Woo00a, RG91, RG93], Wooldridge e Rao e Georgeff fazem uma divisão das fórmulas em dois tipos: fórmulas de estado e fórmulas de caminho. Wooldridge apresenta as fórmulas de caminho como sendo aquelas construídas a partir de conectivos temporais e as de estados são aquelas sem operadores temporais ou que são prefixadas com A ou E.

Em se tratando de componentes temporais, deve-se ter o cuidado com a quantificação em contextos temporais. Por exemplo $\Box\exists x \centerdot PL(x)$ e $\exists x \centerdot \Box PL(x)$ são fbf onde sendo "PL" = professor de lógica, então, a primeira afirma que sempre irá existir um x tal que este será professor de lógica; enquanto a segunda afirma que existe um determinado x o qual será para sempre professor de lógica.

4.1.3 Componente de Ações

Após cada estado de tempo um(a) evento (ação) é esperado(a) e é ele(a) que vai ocasionar ou não mudanças das propriedades válidas no próximo estado de tempo após tal evento (ação). As ações são indicadas pelas letras gregas $\alpha\_$ e $\beta\_$ e os eventos, denotados por $e\_$, podem ser classificados como primitivos – constituídos por apenas uma ação – ou não primitivos – constituídos por mais de uma ação através dos construtores de ações explicitados na tabela 4.4. A partir desses construtores é possível construir qualquer tipo de controle de fluxo da programação. Por exemplo[2]:

While φ **do** $\alpha \equiv\ [(\varphi?;\alpha)/(\neg\varphi?)]*$

Repeat α **until** $\varphi \equiv\ \alpha;[\neg\varphi?;\alpha)/\varphi?]*$

Fórmula	**Interpretação**
$\alpha;\alpha'$	α é seguido de α'
$\alpha\|\alpha'$	Ou exclusivo entre α e α'
$\alpha*$	Ação α ocorre repetidas vezes
$\varphi?$	φ é satisfeito

Tabela 4.4: Construtores de ações.

Observação 4.1.2 *Os construtores sobre ações da tabela 4.4 podem também ser definidos sobre eventos (sejam eles primitivos ou não), denotados por e. Pois, as ações (denotadas pelas letras gregas α e β) são, na verdade, eventos primitivos.*

Em [Woo00a] são definidos os operadores explicitados na tabela 4.5.

[2]Exemplos retirados de [Woo00a].

Fórmula	**Interpretação**
$Happens\,\alpha$	α acontece a seguir
$Achvs\,\alpha\,\varphi$	α ocorre e torna φ satisfeita
$Agts\,\alpha,\ g$	Grupo g é requerido para realizar α

Tabela 4.5: Operadores sobre eventos.

4.1.4 Componente BDI

Este componente incorpora os conectivos modais Bel, Des e Int à lógica de predicados. Com isso, sentenças da seguinte forma são permitidas: $\exists i \,\textbf{.}\, (Bel_i\text{ENSOLARADA}(natal))$, cuja tradução seria "existe pelo menos um determinado agente i o qual acredita que Natal (uma constante) é ensolarada (predicado unário); e $\forall i, j \,\textbf{.}\, \text{MÃE\_DE}(i, j) \rightarrow (Bel_i\, \text{LINDO}(j))$, cuja tradução é "todas as mães acham seu(s) filho(s) lindo(s)". A interpretação de tais conectivos modais será explicitada na sub-seção 4.1.5.

Wooldridge define o domínio $\mathbf{D} = \langle D_{Ag}, D_{Ac}, D_{Gr}, D_U \rangle$, onde D_{Ag} é o domínio dos agentes, D_{Ac} é o domínio das ações, D_{Gr} é o domínio dos grupos de agentes e D_U é o domínio de coisas (outros indivíduos). Seja W um conjunto de mundos e T um conjunto de instantes de tempo então, em [Woo00a], as relações de acessibilidade são definidas da seguinte forma:

$$B : D_{Ag} \rightarrow \wp\{W \times T \times W\}$$

$$D : D_{Ag} \rightarrow \wp\{W \times T \times W\}$$

$$I : D_{Ag} \rightarrow \wp\{W \times T \times W\}$$

Uma situação é um par (w, t), podendo também ser denotado por w_t e o conjunto de todas as situações é definido por $S_w = \{(w, t) \,/\, w \in W \text{ e } t \in T_w\}$ onde W o conjunto de todos os mundos de um modelo de Kripke e T_w o conjunto de todos os instantes de tempo do mundo w. Segundo Wooldridge ([Woo00a]), B é definido como uma função tal que $B(i)$ mapeia uma situação (um par (w, t)) aos mundos de crença onde a crença de um dado agente é verdadeira. Isto é, se houver a fórmula $Bel_i(\varphi)$ em (w, t)[3], logo haverá mundos de crença w' relacionado a w_t no qual φ é satisfeito em w' no tempo t. Wooldridge também define $B_t^w(i) = \{w' \,/\, \langle w, t, w' \rangle \in B(i)\}$[4] como sendo o conjunto de todos os mundos de crença de um dado agente i originados a partir da situação (w, t).[5]

Como D e I são definidos similarmente à B. Então defini-se $Des_i(\varphi)$ e $Int_i(\varphi)$ analogamente a $Bel_i(\varphi)$ e $D_t^w(i)$ e $I_t^w(i)$ analogamente a $B_t^w(i)$.

B é uma relação serial, transitiva e euclidiana, ou seja: para cada situação (w, t) existe pelo menos algum w' que se relaciona com tal situação via relação B (serialidade); se wBw' e $w'Bw''$ então wBw'' (transitividade); e se wBw' e wBw'' então $w''Bw'$ (propriedade euclidiana).

[3] Aqui comete-se um abuso de linguagem uma vez que uma fórmula não se encontra em um mundo. Na verdade, ela recebe o valor verdadeiro neste mundo.

[4] Nota-se que não se faz o mapeamento dos símbolos sintáticos aos domínios de **D**, por isso é necessário explicitar que em $Bel_i(\varphi)$, i representa um agente, enquanto em $B_t^w(i)$, i é realmente o agente.

[5] $B_t^w(i)$ é na verdade a terceira projeção de $B(i)$, ou seja, $\pi^3(B(i))$.

Observação 4.1.3 *Note que para todas as propriedades explicitadas anteriormente, os instantes de tempo t não podem ser distintos. Por exemplo: pela propriedade transitiva tem-se que se $w_{t1}Bw'$ e $w'_{t2}Bw''$ então há de haver um mundo $t1 \in w$ $w_{t1}Bw''$. Entretanto, não tem como assegurar que haja esse instante de tempo $t1 \in w$. Isto se repete para a propriedade euclidiana.*

D e I são relações seriais.

Em se tratando do componente BDI, lembra-se que os operadores modais Bel, Des e Int operam sobre "contextos opacos". A noção de opacidade indica que a lógica BDI não se configura como uma lógica extensional de Frege, ou seja, ela é um tipo de lógica cujas expressões são dependentes apenas das denotações de suas sub-expressões e estas não podem ser substituídas por valores equivalentes ou idênticos.

4.1.5 Sobre a Sintaxe e a Semântica de LORA

LORA é uma lógica multi-tipos, suas constantes e variáveis podem ser do domínio dos agentes, ações, grupos ou "coisas" (outros indivíduos). A sua linguagem é definida por um conjunto enumerável de predicados, variáveis e constantes de cada tipo citado, de símbolos de pontuação e de alguns dos símbolos lógicos, citados na subseção anterior onde outros símbolos são definidos por derivação.

A semântica de LORA é uma semântica de Kripke, ou seja, um modelo $M = \langle T, R, W,$ $\mathbf{D}, Act, Agt, B, D, I, C, \Phi\rangle$, onde T é um conjunto de instantes de tempos; $R \subseteq T \times T$ é uma relação entre esses instantes, $\mathbf{D} = \langle D_{Ag}, D_{Ac}, D_{Gr}, D_U\rangle$; $Act : R \to D_{Ac}$ associa uma ação a um arco R; B, D e I são as relações de acessibilidade; C é uma função que interpreta constantes; e Φ é uma função que interpreta predicados. As interpretações das fórmulas são dadas em duas partes: interpretação das fórmulas de caminho e interpretação das fórmulas de estado, conforme é mostrado, respectivamente, pelas tabelas 4.6 e 4.7.

Fórmula	**Interpretação**
$M, V, w, (w_{t0}, w_{t1}, \ldots) \models_c \varphi$	sse $M, w, t0 \models_e \varphi$, $M, w, t1 \models_e \varphi$,...
$M, V, w, c \models_c \neg\varphi$	sse $M, w, c \not\models_c \varphi$
$M, V, w, c \models_c \varphi \vee \psi$	sse $M, V, w, c \models_c \varphi$ ou $M, V, w, c \models_c \psi$
$M, V, w, c \models_c \forall x \cdot (\varphi)$	sse $M, V \cdot \{x \mapsto d\}, w, c \models_c \varphi$ para todos os indivíduos x do mesmo tipo
$M, V, w, c \models_c \varphi \mathcal{U} \psi$	sse $\exists u \in \mathbb{N}/M, w, c(u) \models_c \psi$ e $\forall v \in \mathbb{N} \,\centerdot\, ((0 \leqslant v < u) \to M, w, c(v) \models_c \varphi)$
$M, V, w, c \models_c \bigcirc\varphi$	sse $M, V, w, c(1) \models_c \varphi$
$M, V, w, c \models_c (Happens(\alpha))$	sse $\exists n \in \mathbb{N} \mid occur(\alpha, c, t0, tn)$

Tabela 4.6: Semântica de Caminhos da LORA.

Fórmula	**Interpretação**
$M, w, t \models_e P(\tau_1, \ldots, \tau_n)$	sse $(V_\rho(\tau_1), \ldots, V_\rho(\tau_n)) \in \Phi(P, t)$, sendo $\tau_j \in \Sigma_C$ ou $\tau_j \in X, \forall j.(1 \leq j \leq n)$
$M, V, w, t \models_e \neg\varphi$	sse $M, V, w, t \not\models_e \varphi$
$M, V, w, t \models_e \varphi \vee \psi$	sse $M, V, w, t \models_e \varphi$ ou $M, V, w, t \models_e \psi$
$M, V, w, t \models_e \forall x \cdot \varphi$	sse $M, V \cdot \{x\| \rightarrowtail d\}, w, t \models_e \varphi$ para todos os indivíduos x do mesmo tipo
$M, V, w, t \models_e (Bel_i(\varphi))$	sse $\forall w' \in W \centerdot (w' \in B_t^w(i) \rightarrow M, V, w', t \models \varphi)$
$M, V, w, t \models_e (Des_i(\varphi))$	sse $\forall w' \in W \centerdot (w' \in D_t^w(i) \rightarrow M, V, w', t \models \varphi)$
$M, V, w, t \models (Int_i(\varphi))$	sse $\forall w' \in W \centerdot (w' \in I_t^w(i) \rightarrow M, V, w', t \models \varphi)$
$M, V, w, t \models_e Agts\, \alpha\, g$	sse $Agts(\alpha, t) = g$
$M, V, w, t \models_e (\tau_1 = \tau_n)$	sse $V_{\rho^t}^w(\tau_1) = V_{\rho^t}^w(\tau_n)$
$M, V, w, t \models_e (i \in g)$	sse $\|i\| \in \|g\|$
$M, V, w, t \models_e A\varphi$	sse $\forall c \in CAMINHOS(w) \centerdot (c(0) = t \rightarrow M, V, w, t \models_c \varphi)$

Tabela 4.7: Semântica de Estados da LoRA.

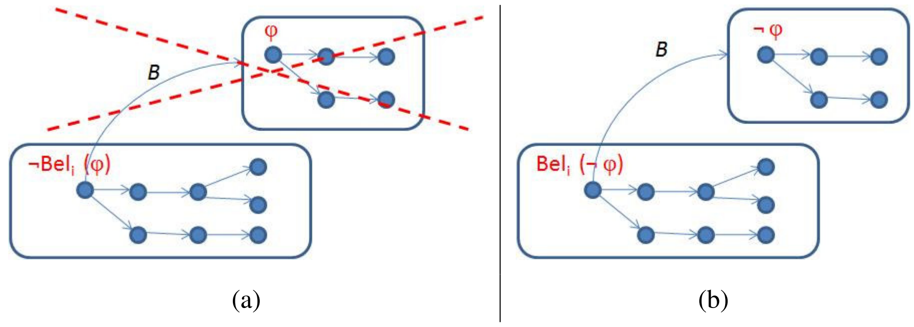
Observação 4.1.4

c é um caminho o qual pode ser representado por $(w_{t0}, w_{t1}, \ldots)$ e se φ é uma fórmula de estado, então, $c(0)$ indica o primeiro instante de tempo (0) do caminho c; e $CAMINHOS(w)$ é interpretado como todos os caminhos da estrutura temporal alocada no mundo w. E $Happens(e)$ é definido através do predicado $occur(e, c, 0, u)$ onde e é um evento que acontece no caminho c mais exatamente entre os dois tempos (de origem e destino) 0 e u.

V em LoRA *é uma função de atribuição de variáveis que associa estas a elementos do domínio: V : $Var \rightarrow$**D**.*

Note que $M, V, w, t \models_e \neg\varphi$ (verdadeiro apenas na situação (w, t)) é diferente de $M, V, w, c \models_c \neg\varphi$ (verdadeiro em todo o caminho c). Entretanto, há um exagero de notação, uma vez que as letras c e w, t, prévias a $\models$ já indicam onde a fórmula é satisfatível.

Note que $M, V, w, t \models \neg Bel(\varphi)$ é diferente de $M, V, w, t \models Bel(\neg\varphi)$; bem como, há distinção semântica entre as fórmulas $\neg Des(\varphi)$ e $Des(\neg\varphi)$; e $\neg Int(\varphi)$ e $Int(\neg\varphi)$. Não acreditar em algo ($\neg Bel(\varphi)$) não significa que ele acredita que o contrário é o caso ($Bel(\neg\varphi)$) – perceba que a segunda sentença é bem mais forte. Formalmente, a primeira sentença significa que há um mundo de crença onde φ é falso e a segunda significa que em todos os mundo de crença $\neg\varphi$ é verdadeiro (vide figura 4.3 onde (a) representa o caso da primeira sentença e (b) representa a segunda). Da mesma forma ocorre com os modais Des e Int em que há uma grande diferença entre não desejar ou não pretender realizar algo – influenciando ou controlando a conduta do agente – e tentar (pretender ou desejar) não satisfazer uma propriedade φ. Vista então a diferença entre os dois esquemas de fórmulas, pode surgir a dúvida: ao negar uma fórmula que envolve um dado operador modal $\overline{M}$ cuja relação de acessibilidade correspondente é $R'_{\overline{M}}$, então $\neg(\overline{M}(\varphi))$ seria equivalente a $\neg\overline{M}(\varphi)$ ou a



(a) (b)

Figura 4.3: Casos representativos de $\neg Bel_\imath(\varphi)$ e de $Bel_\imath(\neg\varphi)$.

$\overline{M}(\neg\varphi)$? Verificar-se-á, supondo o caso de $\overline{M} = Bel$. Daí, vem: $\neg(Bel_i(\varphi))$ é traduzido como "não é o caso de i acreditar em φ", ou "não é o caso de i acreditar que fará φ", ou seja, i não acredita em (ou que fará) φ – φ pode ser $Happens(\alpha)$ como "Aconteceu construir uma casa" ou uma fórmula que não indica uma ação tal como "a casa é azul"– enquanto $\neg Bel_i(\varphi)$ é traduzido como "i não acredita que fará φ", ou "i não acredita em φ"; e $Bel_i(\neg\varphi)$ é traduzido para a linguagem natural como "i acredita que não fará φ" ou "i acredita em $\neg\varphi$". Percebe-se portanto que $\neg(Bel_i(\varphi)) \equiv \neg Bel_i(\varphi)$. Da mesma forma se procede para Des e Int.

Por fim, Wooldridge em seu livro prova que se há uma dada inter-relação entre B, D e I, então, há uma dada relação condicional entre fórmulas envolvendo os conectivos modais Bel, Des e Int, como por exemplo: se $I_t^w(i) \cap D_t^w(i)$ então $Des_i(\varphi) \rightarrow Int_i(\varphi)$. O inverso não é provado e é impossível prová-lo[6]. O autor também não escolhe uma dentre toda as inter-relações entre atitudes mentais, ele prefere deixar em aberto para tornar a sua formalização mais genérica, apesar de explicitar qual seria a inter-relação crença-desejo mais aceitável.

4.2 Lógica BDI de Rao e Georgeff

A lógica BDI proposta por Rao e Georgeff em [RG91, RG93] possui uma semântica justificada em alguns pontos pela própria teoria de Bratman. Todavia, tal lógica apresenta algumas discordâncias para com a teoria de ações de agentes racionais proposta em [Bra87], bem como, algumas discordâncias entre suas abordagens semântica e formal.

4.2.1 Sintaxe

O alfabeto de símbolos lógicos é composto por:

Operadores de 1ª Ordem: $\vee$, $\neg$, $\rightarrow$, $\exists$.

Operadores Temporais: $\bigcirc$, $\Diamond$, $\Box$, $\mathcal{U}$.

Operadores de Ações: $succeeds$, $fails$, $does$, $succeeded$, $failed$, $done$.

[6]Tal ponto será discutido com mais detalhes na sub-seção 4.4.

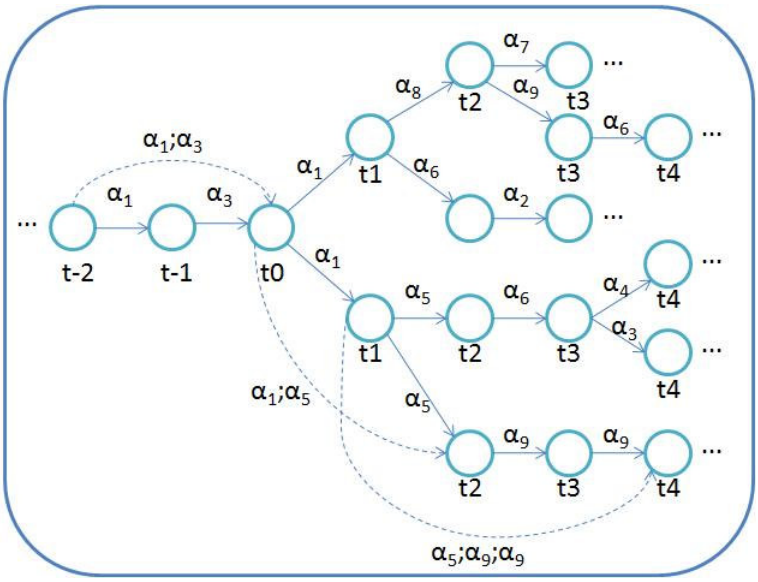


Figura 4.4: Exemplo de um mundo segundo Rao e Georgeff.

Operadores modais e BDI: $optional$, $inevitable$, Bel, $Goal$, Int.[7].

Observação 4.2.1 *Note que não há um operador sobre grupos de agentes. E também não há identificação sobre quem é o responsável pela atitude mental, demonstrando assim que tal formalização é, possivelmente, capaz de representar o raciocínio de um agente, mas não é suficiente para representar um sistema multi-agente e modelar o raciocínio de diversos agentes.*

4.2.2 Semânticas Informais da Lógica de Rao e Georgeff

De forma similar à LORA, Rao e Georgeff definem uma semântica de Kripke composta por mundos possíveis onde cada mundo é uma árvore temporal ("*time-tree*") e um ponto (estado) particular desse mundo é um par (w, t). Cada um desses estados temporais possuem um passado linear e um futuro ramificado e as relações entre esses estados são também totais e transitivas onde cada arco é relacionado a um evento (a figura 4.4 mostra um exemplo de um mundo definido por Rao e Georgeff).

Rao e Georgeff usam em sua lógica BDI, as atitudes mentais de crença, objetivo e intenção. Assim como Levesque e Cohen em [CL87], a substituição de desejo por objetivo se dá devido a Rao e Georgeff considerarem que em qualquer estado de tempo os agentes podem ter desejos conflitantes ou até mesmo contraditórios[8]. Então, Rao e Georgeff definem a nova atitude mental objetivo como sendo um conjunto de desejos não-conflitantes[9].

Semelhantemente, a Cohen e Levesque ([CL87]), que definem a propriedade de Realismo ("o agente adota apenas objetivos os quais ele acredita serem exequíveis"), Rao e Georgeff definem a propriedade Realismo Forte: o agente acredita que ele pode opcionalmente realizar seus objetivos através da escolha dos eventos

[7]Onde $optional$ e $inevitable$ são os quantificadores de caminhos cujas interpretações são similares à A e E da LORA respectivamente

[8]Esta concepção também é defendida por Bratman indicando que o conceito de desejo é semelhante nos trabalhos [RG91, RG93, CL87]

[9]Cohen e Levesque tratam o problema de operar desejos conflitantes de forma semelhante. Eles assumem que quando o agente vai agir, ele deve escolher um conjunto específico de desejos não-conflitantes (objetivos)

a serem executados. Esta noção de realismo forte é forçada através da adoção da compatibilidade entre objetivos e intenções definida como:

> Para cada mundo w acessado via relação de crença B em um dado momento t, existe um mundo w' acessado via relação de objetivo que é um sub-mundo de w neste tempo t.

Rao e Georgeff fazem 3 definições importantes em ambos os artigos em que são definidas teorias formais BDI ([RG91, RG93]). São definidos modelos (item 1(a) e 1(b)), mundos (item 2) e sub-mundos (item 3):

1. (a) A semântica de Kripke é um modelo $M = \langle W, E, T, R, U, B, G, I, \Phi \rangle$. Onde W é um conjunto de mundos possíveis, E é um conjunto de eventos, T é um conjunto de instantes de tempo, R é uma relação (arco) entre esses instantes, U é um universo de discurso, B, G e I são as relações de acessibilidade entre mundos (elas ligam uma situação a um mundo de crenças, de objetivos ou de intenções respectivamente). E Φ mapeia entidades de primeira ordem em elementos de U.[10]

   (b) A semântica de Kripke é um modelo $M = \langle W, S, E, T, U, B, G, I, \Phi \rangle$. Onde W é um conjunto de mundos possíveis, S é um conjunto de situações, E é um conjunto de eventos, T é um conjunto de instantes de tempo, U é um universo de discurso, B, G e I são as relações de acessibilidade que mapeiam as situações entre mundos (elas ligam uma situação a um mundo de crenças, de objetivos ou de intenções, respectivamente). E Φ mapeia entidades de primeira ordem em elementos de U.[11]

2. Cada mundo é definido como uma tupla $w = \langle T_w, A_w, S_w, F_w \rangle$ onde $T_w \subseteq T$ é um conjunto de instantes de tempo adjacentes de w, $A_w \subseteq R$, $S_w : T_w \times T_w \to E$ e $F_w : T_w \times T_w \to E$ são funções que mapeiam instantes de tempos adjacentes ao conjunto de eventos E.

3. Um mundo w' é submundo de w, denotado por $w' \sqsubseteq w$, sse w' é uma sub-árvore de w com as mesmas valorações para as sentenças, ou seja, $w' \sqsubseteq w$ sse (a) $T_{w'} \subseteq T_w$ (b) para todo $u \in T_{w'}$, $\Phi(q, w', u) = \Phi(q, w, u)$, onde q é um símbolo de predicado, (c) para todo $u \in T_{w'}$, $R_u^w = R_u^{w'}$ e (d) $A_{w'}$ é igual a A_w, quando este é restrito aos estados de $T_{w'}$ e similarmente para $S_{w'}$ e $F_{w'}$. Diz-se ainda que w' é um sub-mundo estrito de w sse $w' \sqsubseteq w$ e $w \not\sqsubseteq w'$.

Observação 4.2.2 *Percebe-se que há a mesma definição de semântica de mundos possíveis entre os artigos explicitados acima. Quanto a definição de sub-mundo; ela não é muito clara. Nesta definição é dada à expressão sub-árvore, um sentido lógico não usual. E se-interpreta $w' \sqsubseteq w$ sse w é uma extensão (no sentido algébrico) de w' onde todas as sentenças, relações de acessibilidade e relações entre instantes de tempo, ações, sucessos e falhas que se encontram no mundo w' são também encontrados da mesma forma em w. Por fim, em [RG91, RG93], é requerido que R seja total, transitiva e linear no passado para forçar que um dado instante de tempo tenha um passado linear e um futuro ramificado.*

As fórmulas desta lógica também são divididas em fórmulas de estado e fórmulas de caminho. As fórmulas do 1º tipo são apresentadas como aquelas as quais são valoradas em um dado instante de tempo de

[10] Parafraseado de [RG91].
[11] Parafraseado de [RG93].

um dado mundo e as do 2º tipo são aquelas fórmulas valoradas em um dado caminho da estrutura temporal de um dado mundo. Em outras palavras, tem-se:

- Qualquer fórmula de 1ª ordem é uma fórmula de estado.
- Seja ϕ_1 e ϕ_2 fórmulas de estado e x é um objeto ou um evento então $\neg\phi_1$, $\phi_1 \vee \phi_2$, $\phi_1 \mathcal{U} \phi_2$, $\Box\phi_1$, $\Diamond\phi_1$, $\bigcirc\phi_1$, $\exists\phi_1(x)$ são fórmulas de estado.
- Seja 'e' um evento então $Succeeds(e)$, $Fails(e)$, $does(e)$, $Succeeded(e)$, $Failed(e)$ e $done(e)$ são fórmulas de estado.
- Seja ϕ uma fórmula de estado então $inevitable(\phi)$ e $optional(\phi)$ são fórmulas de caminho.
- Seja ϕ uma fórmula de estado então $Bel_i(\phi)$, $Goal_i(\phi)$ e $Int_i(\phi)$ são fórmulas de estado.
- Toda fórmula de estado é também fórmula de caminho.
- Seja ϕ_1 e ϕ_2 fórmulas de caminho, então $\neg\phi_1$, $\phi_1 \vee \phi_2$, $\phi_1 \mathcal{U} \phi_2$, $\Box\phi_1$, $\Diamond\phi_1$, $\bigcirc\phi_1$ são fórmulas de caminho.

Os quantificadores sobre caminhos descrevem fácil e claramente as opções avaliáveis para o agente, possibilitando quais são as propriedades que o agente pode escolher satisfazê-las e quais propriedades não estão sob o domínio do agente, pois estas são inevitáveis. Então, é importante diferenciar as fórmulas opcionais (O-fórmulas) das fórmulas inevitáveis (I-fórmulas) (tal como é feito em [RG91, RG93]). As O-fórmulas são fórmulas bem-formadas (fbf) que contém ocorrências positivas de $optional$ (E) ou ocorrências negativas de $inevitable$ e I-fórmulas são fbf que contém ocorrências positivas de $inevitable$ (A). Rao e Georgeff definem $inevitable$ em função de $optional$, ou seja considerando $optional$ como primitivo, e vice-versa, ou seja definem $optional$ em função de $inevitable$, da seguinte maneira:

$$inevitable(\varphi) \stackrel{\text{def}}{=} \neg optional(\neg\varphi) \;\text{ e }\; optional(\varphi) \stackrel{\text{def}}{=} \neg inevitable(\neg\varphi)$$

Caracterização das crenças, objetivos e intenções segundo Rao e Georgeff

A noção de mundos de crenças, desejos ou objetivos e intenções abordada em 4.1.5 é a mesma definida por Rao e Georgeff em[RG91, RG93]. No entanto, nestes artigos é capturada uma inter-relação entre as atitudes mentais na qual todos os mundos de intenções são sub-mundos dos mundos de objetivos que por sua vez são sub-mundos dos mundos de crenças, denotado por $I_t^w \subseteq G_t^w \subseteq B_t^w$. Isto é decorrência da adoção do realismo forte. Logo, tem-se que para todo mundo proveniente de uma relação de crença em um dado tempo t (bw_t) existe um mundo proveniente de uma relação de objetivo, nesse momento t (gw_t) tal que gw_t é um sub-mundo de bw_t $(gw_t \sqsubseteq bw_t)$. Um exemplo desta descrição de submundos é vista na figura 4.5.

A mesma relação de compatibilidade entre crenças e objetivos é exigida entre objetivos e intenções, isto é, para todo mundo proveniente de uma relação de objetivos em um dado tempo t (gw_t) existe um mundo proveniente de uma relação de intenção nesse momento t (iw_t) tal que iw_t é um sub-mundo de dw_t, ou seja

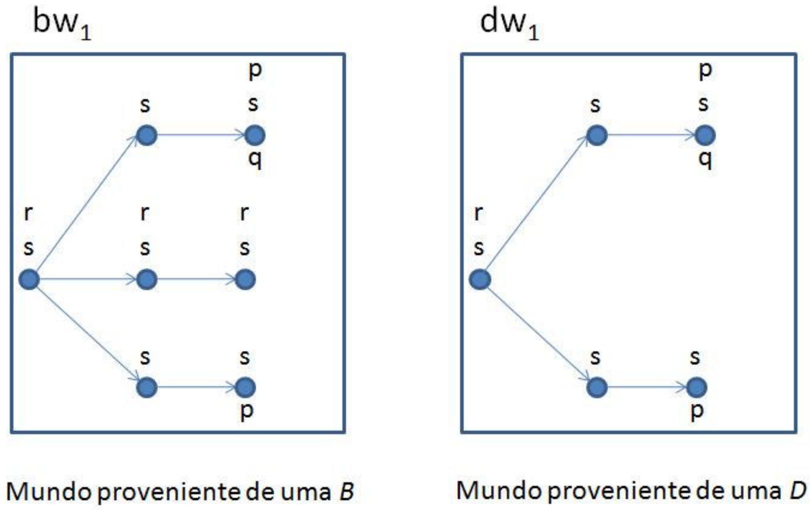


Figura 4.5: Exemplo de um esquema de mundos composto por bw_1 e seu sub-mundo dw_1.

$iw_t \sqsubseteq gw_t$. Isso significa que as intenções, que são similarmente representadas por mundos acessados via relação de intenção, são a parte dos objetivos a qual o agente se compromete em tentar realizar; em outras palavras, o agente só pretende realizar certos cursos de ações caso estes sejam um dos seus objetivos opcionais.

4.2.3 Axiomatização e Condições Semânticas

A axiomatização para crenças é pelo sistema modal S5 (ou KD45) enquanto os axiomas referentes a G e I são K e D (sistema KD), pois os desejos e as intenções devem ser consistentes e fechados sobre a implicação. Os outros axiomas são referentes às inter-relações entre as atitudes mentais e relações entre intenções/crenças e fórmulas de ações. Tais axiomas[12] e suas respectivas condições semânticas (denotadas por SC_) podem ser observadas a seguir:

[12] O conjunto de axiomas apresentados em [RG91] é diferente do apresentado em [RG93]. Nesta subseção apresenta-se os definidos em [RG93].

[A1] $Goal(optional(\varphi)) \rightarrow Bel(optional(\varphi))$

[SC1] $\forall w' \in B_t^w \exists w'' \in G_t^w$ t.q. $w'' \sqsubseteq w'$

[A2] $Int(optional(\varphi)) \rightarrow Goal(optional(\varphi))$

[SC2] $\forall w' \in G_t^w \exists w'' \in I_t^w$ t.q. $w'' \sqsubseteq w'$

[A3] $Int(\varphi) \rightarrow Bel(Int(\varphi))$

[SC3] $\forall w' \in B_t^w$ e $\forall w'' \in I_t^w$ tem-se que $w'' \in B_t^w$

[A4] $Goal(\varphi) \rightarrow Bel(Goal(\varphi))$

[SC4] $\forall w' \in B_t^w$ e $\forall w'' \in G_t^w$ tem-se que $w'' \in B_t^w$

[A5] $Int(\varphi) \rightarrow Goal(Int(\varphi))$

[SC5] $\forall w' \in G_t^w$ e $\forall w'' \in I_t^w$ tem-se que $w'' \in G_t^w$

[A6] $Int(inevitable(\bigcirc done(e))) \rightarrow inevitable(\bigcirc done(e))$

[SC6] Se o agente tem a intenção de realizar uma ação primitiva determinística então ele a fará.

[A7] $done(e) \rightarrow Bel(done(e))$

[SC7] O agente deve estar ciente de todos os eventos primitivos correntes no mundo.

Esses axiomas compõem o sistema base de axiomas o qual pode ser estendido através da adição de algum(uns) dos três axiomas referentes às estratégias de compromisso dos agentes.

Axioma referente à estratégia de compromisso cego : Neste tipo de compromisso, se o agente pretende inevitavelmente realizar ou satisfazer φ então ele manterá sua intenção até que ele acredite que a realizou/satisfez.

A10a $(Int(inevitable(\Diamond\varphi))) \rightarrow (inevitable(Int(inevitable(\Diamond\varphi \,\mathcal{U}\, Bel(\varphi))))).$

Os axiomas referentes às estratégias de compromisso podem referir-se ao compromisso com meios (caso φ seja uma fórmula de ações) e com fins (caso φ seja uma condição que deve ser verdadeira no futuro). Note que eles tratam apenas o compromisso a meios e fins os quais são inevitáveis, ou seja, nada é falado com respeito ao compromisso do agente realizar opcionalmente meios e fins.

Axioma referente à estratégia de compromisso simples : Nesta estratégia, o agente mantém suas intenções até que ele acredite que a satisfez ou até não acreditar que seja possível satisfazê-la.

A10b $Int(inevitable(\Diamond\varphi)) \rightarrow$
$inevitable(Int(inevitable(\Diamond\varphi)))\mathcal{U}\,(Bel(\varphi) \vee (\neg Bel(optional(\Diamond\varphi)))).$

Axioma referente à estratégia de compromisso mente-aberta : Neste tipo de compromisso, a intenção é mantida até que o agente acredite que a satisfez ou enquanto ele a deseje satisfazê-la. E é formalizado pelo seguinte axioma:

A10c $Int(inevitable(\Diamond\varphi)) \rightarrow$
$inevitable(Int(inevitable(\Diamond\varphi)))\,\mathcal{U}\,(Bel(\varphi) \vee (\neg Goal(optional(\Diamond\varphi))))$

Outras formas de compromisso podem ser obtidos pela mistura entre tais compromissos; em [RG91], por exemplo, cita-se a possibilidade do agente ser mente-aberta com relação aos seus fins, mas ter um compromisso simples com relação aos meios referentes aos fins; se tornando interessante pois quanto mais longe de realizar suas pretensões mais fácil de abandoná-las ou substituí-las.

Há ainda dois axiomas que podem ser adicionados; caso se deseje formalizar um agente competente (**A8** $Bel(\varphi) \rightarrow \varphi$) e um agente confiante (**A9** $done(e) \rightarrow succeeded(e)$). Contudo, tais axiomas não serão comentados já que estão fora do escopo deste trabalho definir tais tipos de agentes.

4.2.4 Conclusão e Comparações

Como conclusão do artigo [RG93], os autores afirmam que desenvolveram uma lógica BDI compatível com a teoria de Bratman. Eles vêem esta como uma formalização lógica de intenção e uma teoria filosófica de intenção onde em ambas as intenções são pró-atitudes controladoras de condutas, os desejos são pró-atitudes influenciadoras de condutas e as intenções não podem ser reduzidas em função de desejos e crenças; além de que Bratman (em [Bra87]) distingue os planos quanto estruturas abstratas (bibliotecas de planos) e quanto estados mentais, o que são na essência uma intenção de realizar uma sequência de ações tal como é formalizado por Rao e Georgeff em [RG91, RG93]. Dizem também que, pela primeira vez, uma formalização lógica capturou os processos de revisão de crenças, desejos e intenções apresentados por Allen em [All90]. Apesar disso, admitem que sua formalização encontra-se incompleta com respeito à deliberação e à reconsideração de intenções especificada por Bratman.

Rao e Georgeff, ainda em sua conclusão, afirmam que demonstraram ter solucionado um dos grandes problemas da formalização lógica de modelos BDI: o problema de "efeitos indesejados" (*"side-effects"*). Este problema se dá pela dificuldade de satisfazer as sete propriedades necessárias a um agente cognitivo exposto em [Bra87] (apresentadas na seção 2.1), especialmente, a sétima. Esta diz que se um agente pretende satisfazer p e é verdade que q é consequência da realização de p então não é necessário que o agente pretenda q. Tal propriedade pode ser formalizada por: $Int(\varphi) \wedge (\varphi \rightarrow \phi) \rightarrow \neg\Box Int(\phi)$. Enquanto o problema de efeitos indesejados pode ser expresso por $Int(\varphi) \wedge Bel(\varphi \rightarrow \phi) \wedge \neg Int(\phi)$

Um exemplo de efeito indesejado é exposto em [RG91] (vide figura 4.6), onde supõe-se que um agente acredita que é inevitável que se ele for ao dentista (α_1) arrancar um dente, isto implicará nele sentir dor (mundo b1), no entanto, ele deseja arrancar o dente e não deseja que arrancar o dente implique nele sentir dor (mundo d2) e tem a intenção de ir ao dentista e mesmo assim arrancar o dente e não sentir dor (mundo $i2$). É como um exemplo citado anteriormente que mesmo o agente crendo que comer demasiadamente o tornará gordo, ele deseja comer e não obrigatoriamente deseje tornar-se gordo.

Os autores afirmam ter contornado o problema já que em sua formalização (axioma A1) é possível ocorrer $Bel(optional(\varphi))$ e $\neg Goal(optional(\varphi))$ (caso, no exemplo, $\varphi = p \rightarrow q$ o problema de efeito indesejado apresentado pela figura poderia ser resolvido). Porém, de acordo com SC1, ou seja, de acordo com a formalização dada ao realismo forte, não é possível ocorrer $Bel(\varphi)$ e $\neg Goal(\varphi)$. Esta divergência e algumas outras serão estudadas minuciosamente na próxima seção.

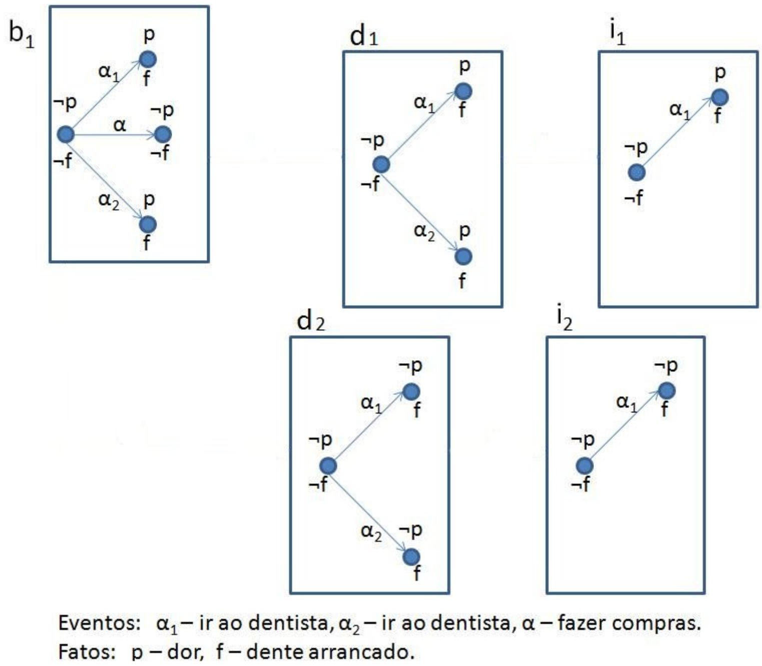


Figura 4.6: Exemplo de mundos possíveis provenientes das relações B, D e I.

Observação 4.2.3 *Na figura 4.6 é possível avaliar também que nos mundos de intenções i_1 e i_2 no último estado de tempo, há a fórmula $Succeeded(\alpha_1)$ significando que o agente pretende realizar α_1 com sucesso pois ele tem este objetivo ($i_1 \sqsubset d_1 \sqsubset b_1$ e $i_2 \sqsubset d_2$). Porém, o agente não pode garantir o sucesso de α_1, pois isso dependerá do ambiente no qual o agente se encontra, ou seja, apesar de $Succeeded(\alpha_1)$ ser verdadeira em todos os i_n, não necessariamente ela será verdadeira no mundo "real" w em que o agente se encontra. Apesar de que, como $inevitable(\Diamond f)$ é válido em todos os d_n, então $Bel_i\,(optional(\Diamond f))$, ou seja: o agente i crê que f é exequível.*

4.3 Relações entre as Atitude Mentais

Nesta seção serão tratados os tipos de relações que ocorrem entre as atitudes mentais. Wooldridge, em [Woo00a], apresenta 8 relações; são elas: $=$, $\neq$, $\cap$, $\cap_{sub}$, $\cap_{sup}$, $\subseteq$, $\subseteq_{sub}$ e $\subseteq_{sup}$ (super-conjunto estrutural). Algumas dessas relações são definidas por ele. E algumas destas serão apresentadas[13], de maneira formal, a seguir.

Definição 4.3.1 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente.*

1. $R'(\imath) = \overline{R'}(\imath)$ *sse* $R'(\imath) \subseteq \overline{R'}(\imath)$ *e* $\overline{R'}(\imath) \subseteq R'(\imath)$

2. $R'^{\,w}_{\,t}(\imath) \cap_{sub} \overline{R'}^{w}_{t}(\imath) \stackrel{\text{def}}{=} \{w' \,/\, w' \in \overline{R'}^{w}_{t}(\imath)$ *e* $\exists w'' \in R'^{\,w}_{\,t}(\imath)$ *t.q.* $w'' \sqsubseteq w'\}$

[13]A apresentação das definições das relações entre atitudes mentais explicitadas aqui são as versões formais das definições de relações apresentadas por Woodridge em [Woo00a].

3. $R'^{w}_{t}(\imath) \cap_{sup} \overline{R'}^{w}_{t}(\imath) \stackrel{\text{def}}{=} \{w' \,/\, w' \in \overline{R'}^{w}_{t}(\imath)$ *e* $\exists w'' \in R'^{w}_{t}(\imath)$ *t.q.* $w' \sqsubseteq w''$

4. $R'^{w}_{t}(\imath) \subseteq \overline{R'}^{w}_{t}(\imath)$ *sse* $\forall w', w''.(w' \in R'^{w}_{t}(\imath) \rightarrow (w'' \in \overline{R'}w_t(\imath)) \wedge CAMINHOS(w') \subseteq CAMINHOS(w''))$

5. $R'^{w}_{t}(\imath) \subseteq_{sub} \overline{R'}^{w}_{t}(\imath)$ *sse* $\forall w'.(w' \in R'^{w}_{t}(\imath) \rightarrow \exists w''.(w'' \in \overline{R'}w_t(\imath)) \wedge w' \sqsubseteq w'')$

6. $R'^{w}_{t}(\imath) \subseteq_{sup} \overline{R'}^{w}_{t}(\imath)$ *sse* $\forall w'.(w' \in R'^{w}_{t}(\imath) \rightarrow \exists w''.(w'' \in \overline{R'}w_t(\imath) \wedge w'' \sqsubseteq w'))$

Detalhando um pouco mais a inter-relação $\cap$, seja ela $\cap_{sup}$ ou $\cap_{sub}$, tem-se que: seja R e $\overline{R}$ relações de acessibilidade B, G ou I, tal que $R \neq \overline{R}$; então $R^{w}_{t}(i) \cap_{sub} \overline{R}^{w}_{t}(i) \neq \emptyset$ ou $\overline{R}^{w}_{t}(i) \cap_{sup} R^{w}_{t}(i) \neq \emptyset$ quer dizer que há um conjunto de mundos provenientes de R (ou simplesmente mundos de R) que são sub-mundos do(s) mundo(s) proveniente(s) de $\overline{R}$. Isto por sua vez implica em:

- Há mundos de $\overline{R}$ que <u>não</u> têm mundos de R como seus sub-mundos
- Há mundos de $\overline{R}$ que têm mundos de R como seus sub-mundos
- Há mundos de R que <u>não</u> são sub-mundos dos mundos provenientes de R.

Em outras palavras, $\overline{R}^{w}_{t}(i) \cap_{sup} R^{w}_{t}(i)$ denota o conjunto dos mundos $w' \in \overline{R}$ para os quais existe algum mundo $w'' \in R$ tal que $w' \sqsubseteq w''$. E $\overline{R}^{w}_{t}(i) \cap_{sub} R^{w}_{t}(i)$ denota o conjunto dos mundos $w' \in \overline{R}$ para os quais existe algum mundo $w'' \in R$ tal que $w'' \sqsubseteq w'$.

Observação 4.3.1 *É impossível que não existam mundos de R ou de $\overline{R}$, pois B, G e I são relações seriais.*

Considerando as semânticas de mundos possíveis propostas por Wooldridge ([Woo00a]) e por Rao e Georgeff ([RG91, RG93]), as seguintes proposições são verdadeiras[14]:

Proposição 4.3.1 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se $\overline{R}^{w}_{t}(i) \subseteq R^{w}_{t}(i)$, então, $\mathcal{M}_i(\varphi) \rightarrow \overline{\mathcal{M}}_i(\varphi)$.*

Proposição 4.3.2 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$, e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se $R^{w}_{t}(i) \subseteq_{sub} \overline{R}^{w}_{t}(i)$, então, para qualquer I-fórmula $inevitable(\varphi)$, $\overline{\mathcal{M}}_i(inevitable(\varphi)) \rightarrow \mathcal{M}_i(inevitable(\varphi))$.*

Proposição 4.3.3 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se $R^{w}_{t}(i) \subseteq_{sup} \overline{R}^{w}_{t}(i)$, então, para qualquer O-fórmula $optional(\varphi)$, $\overline{\mathcal{M}}_i(optional(\varphi)) \rightarrow \mathcal{M}_i((optional(\varphi))$.*

Proposição 4.3.4 *Sejam $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se $R^{w}_{t}(i) \cap_{sup} \overline{R}^{w}_{t}(i) \neq \emptyset$, então, para qualquer I-fórmula $inevitable(\varphi)$, $\mathcal{M}_i(inevitable(\varphi)) \rightarrow \neg\overline{\mathcal{M}}_i(\neg inevitable(\varphi))$.*

[14]Tais proposições foram provadas em [Woo00a]

Proposição 4.3.5 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se $R_t^w(i) \cap_{sub} \overline{R}_t^w(i) \neq \emptyset$, então, para qualquer O-fórmula $optional(\varphi)$, $\mathcal{M}_i(optional(\varphi)) \to \neg\overline{\mathcal{M}}_i(\neg optional(\varphi))$.*

Wooldridge ainda prova que:

- Se $R_t^w(i) \subseteq \overline{R}_t^w(i)$, então, $\mathcal{M}_i(\varphi) \to \overline{\mathcal{M}}_i(\varphi)$.
- Se $R_t^w(i) \cap \overline{R}_t^w(i) \neq \emptyset$, então, $\mathcal{M}_i(\varphi) \to \neg\overline{\mathcal{M}}_i(\neg\varphi)$.
- Se $R_t^w(i) = \overline{R}_t^w(i) \neq \emptyset$, então, $\mathcal{M}_i(\varphi) \leftrightarrow \overline{\mathcal{M}}_i(\varphi)$.

Nas lógicas apresentadas neste trabalho serão usadas as relações '$\cap_{sub}$' ou '$\cap_{sup}$', e, '$\subseteq_{sub}$' ou '$\subseteq_{sup}$'. Na próxima seção, analisar-se-á as teorias formais de Rao e Georgeff e a abordagem semântica da lógica BDI de Wooldridge.

4.4 Problemas nas Formalizações BDI

Apresentou-se neste capítulo as formalizações BDI, baseadas unicamente na teoria de Bratman, mais conhecidas. Uma, proposta por Rao e Georgeff em 1991 ([RG91]) e, posteriormente, revisada e apresentada em uma nova versão em 1993 ([RG93]) e outra, proposta por Wooldridge em 2000 ([Woo00a]) a qual parece ter recebido fortes influências dos trabalhos anteriores. Contudo, percebe-se algumas diferenças: (a) alfabeto dos símbolos lógicos (principalmente com respeito aos operadores de ações); (b) os primeiros apresentam uma teoria formal onde há a escolha explícita de um modelo de agentes BDI o qual obedece ao realismo forte. Enquanto o segundo não apresenta qualquer abordagem formal, ao invés disso, mostra como poder-se-ia obter os diferentes modelos de agentes cognitivos em função das variações do Realismo; (c) o último é capaz de representar um sistema multiagente enquanto os primeiros não; (d) Wooldridge prova que o problema de efeitos indesejados não é solucionado quando o modelo de agente BDI obedece ao realismo forte e afirma que este modelo não é o mais aceitável para a teoria de Bratman enquanto Rao e Georgeff afirmam e "provam" que solucionaram tal problema.

As diferenças mostradas nos itens 'a', 'b' e 'c' refletem na maneira e capacidade de expressão das sentenças em cada uma das formalizações. Entretanto, as discordâncias do item 'd' mostram conclusões contraditórias sobre as mesmas premissas. Aqui serão analisadas tais formalizações.

Rao e Georgeff definem uma semântica informal e uma teoria formal baseada no Realismo forte. Sua teoria formal tem está propriedade assegurada através do axioma A1 (referenciado novamente em 4.1).

$$Goal(optional(\varphi)) \to Bel(optional(\varphi)) \tag{4.1}$$

Os autores afirmam que está propriedade é capturada semanticamente pela compatibilidade crença-objetivo requerendo que, para todo mundo proveniente de uma relação de crença w em um dado instante t, existe um mundo proveniente de uma relação de objetivo tal que este é sub-mundo de w. Rao e Georgeff, em [RG91], apresentam a condição semântica que supostamente expressaria a compatibilidade sintática entre crenças e objetivos (A1) como $\forall w' \in B_t^w$, $\exists w'' \in G_t^w$ tal que $w'' \sqsubseteq w'$. E em [RG93], definem esta condição semântica como:

$$\forall t \in T, \text{ se } \exists w' \in W \text{ tal que } w_t B w' \text{ então } \exists w'' \in W \text{ tal que } w_t G w'', \text{ e } w'' \sqsubseteq w'.$$

Tais condições capturariam a noção de realismo forte forçada na teoria formal pelo axioma A1. A compatibilidade crença-objetivo é denotada em [RG91] por $B_t^w \subseteq_{super} G_t^w$ (em [RG93] não há denotação para tal condição semântica).

Já Wooldridge prova (proposição 4.3.3) que se a condição semântica expressa em 4.2, denotado por $B(i) \subseteq_{sup} G(i)$, é verdadeira, então, o axioma A1[15] é satisfeito.

$$\text{Se } w' \in B_t^w, \text{ então, existe algum } w'' \in G_t^w \text{ tal que } w'' \sqsubseteq w'. \tag{4.2}$$

A compatibilidade entre objetivos e intenções de Rao e Georgeff é forçada pelo axioma A2 (referenciada novamente pela implicação 4.3). Ela é análoga à compatibilidade entre crenças e objetivos e mantém a mesma escolha. Em Wooldridge, a condição semântica, que expressa a fórmula 4.3, é a inter-relação 4.4, a qual pode ser denotada por $G(i) \subseteq_{sup} I(i)$:

$$Int_i(optional(\varphi)) \to Goal_i(optional(\varphi)) \tag{4.3}$$

$$\text{Se } w' \in G_t^w \text{ então existe algum } w'' \in I_t^w \text{ tal que } w'' \sqsubseteq w'. \tag{4.4}$$

Conclui-se que as duas formalizações apresentam duas condições semânticas distintas para interpretar a mesma propriedade sintática (axioma A1 – fórmula 4.1). Portanto, verificar-se-á as compatibilidades de crenças-desejos-intenções presentes nos artigos referenciados de Rao e Georgeff e as inter-relações 4.2 e 4.4 apresentadas em [Woo00a]. Contudo, mostrar-se-á, através das proposições 4.4.1 e 4.4.2, que não há bi-implicação entre elas e as propriedades 4.1 e 4.3, tal como é afirmado por Wooldridge em [Woo00a].

Proposição 4.4.1 *a)* $B_t^w(i) \subseteq_{sup} G_t^w(i)$ *é expresso sintaticamente por* $Goal_i(optional(\varphi)) \to Bel_i(optional(\varphi))$, *mas b)* $Goal_i(optional(\varphi)) \to Bel_i(optional(\varphi))$ *não é expresso semanticamente por* $B_t^w(i) \subseteq_{sup} G_t^w(i)$.

PROVA:

a) Prova análoga à prova da proposição 4.3.3.

b) Não é o caso de: se $\models Goal_i(optional(\varphi)) \to Bel_i(optional(\varphi))$ então $B_t^w \subseteq_{sup} G_t^w$; pois nem os operadores modais $Goal$ e Bel, nem as relações de acessibilidade G e B interpretam qualquer tipo de relação de inclusão. Como pode ser visto no contra-exemplo da figura 4.7 que representa uma situação na qual é assegurada por $\models Goal_i(optional(\varphi)) \to Bel_i(optional(\varphi))$ e não é verdadeira se $B_t^w \subseteq_{sup} G_t^w$, pois o mundo proveniente de G não é submundo do mundo proveniente de B.

■

[15] Analisar-se-á as fórmulas de Wooldridge desconsiderando a diferença entre desejos e objetivos para facilitar a explicação e sem prejuízo de entendimento já que ele mesmo não faz essa distinção.

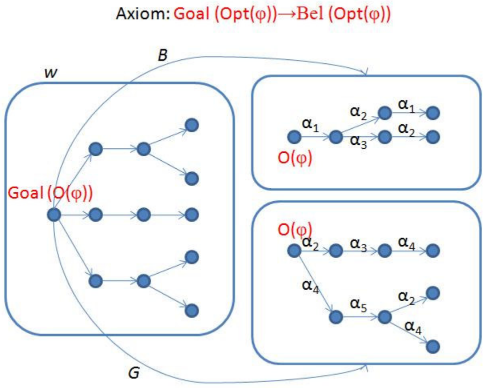


Figura 4.7: Contra-exemplo

Proposição 4.4.2 *a)* $G_t^w(i) \subseteq_{sup} I_t^w(i)$ *é expresso sintaticamente por* $Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$*, mas b)* $Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$ *não é expresso semanticamente por* $G_t^w(i) \subseteq_{sup} I_t^w(i)$*.*

PROVA: Análoga à prova da proposição 4.4.1. ■

Proposição 4.4.3 *a) Se* $\forall w' \in B_t^w, \exists w'' \in G_t^w$ *tal que* $w'' \sqsubseteq w'$*, então,* $\models Goal_i(optional(\varphi)) \rightarrow Bel_i(optional(\varphi))$*. b) Entretanto, se* $\models Goal_i(optional(\varphi)) \rightarrow Bel_i(optional(\varphi))$*, então,* $\forall w' \in B_t^w, \exists w'' \in G_t^w$ *tal que* $w'' \sqsubseteq w'$ *não é o caso.*

PROVA:

a) Suponha que $\forall w' \in B_t^w, \exists w'' \in G_t^w$ tal que $w'' \sqsubseteq w'$ – formalmente tem-se que $\forall w'.(w' \in B_t^w \rightarrow \exists w''.(w'' \in G_t^w \wedge w'' \sqsubseteq w'))$. E suponha $M, V, w_t \not\models Goal(optional(\varphi)) \rightarrow Bel(optional(\varphi))$ para algum $\langle M, V, w, t \rangle$. Daí obtém-se $M, V, w_t \models Goal(optional(\varphi))$ e $M, V, w_t \models \neg Bel(optional(\varphi))$ e pelas semânticas de $Goal$ e de Bel, conclui-se que $\forall w'' \in G_t^w$, $M, V, w_t'' \models optional(\varphi)$ e $\exists w' \in B_t^w$ tal que $M, V, w_t' \not\models optional(\varphi)$. Todavia, pela condição semântica e por $M, V, w_t' \not\models optional(\varphi)$, tem-se que $\exists w'' \in G_t^w$ tal que $M, V, w_t'' \not\models optional(\varphi)$. O que é uma contradição. Logo, a suposição não é verdadeira.

b) Similar à prova da proposição 4.4.1.b. ■

Proposição 4.4.4 *a) Se* $\forall w' \in G_t^w, \exists w'' \in I_t^w$ *tal que* $w'' \sqsubseteq w'$ *então* $Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$*. b) Entretanto se* $Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$ *então* $\forall w' \in G_t^w, \exists w'' \in I_t^w$ *tal que* $w'' \sqsubseteq w'$ *não é o caso.*

PROVA: Prova análoga à prova da proposição 4.4.3. ■

Conforme demonstrado neste capítulo, há ambiguidades entre a teoria formal e a semântica informal da lógica BDI de Rao e Georgeff. Bem como ocorre em LORA, onde há a relação de suficiência entre as propriedades semânticas e as sintáticas (acerca da atitudes mentais), mas tal relação não é necessária, conforme provado nas proposições 4.4.1 e 4.4.2. Assim, pode-se afirmar que Rao e Georgeff desenvolveram uma lógica BDI cujo teorema de corretude não pode ser provado, ou seja, sua teoria formal não corresponde à semântica explicitada; e Wooldridge, caso desenvolvesse uma teoria formal (sejam quais forem as inter-relações entre as atitudes mentais escolhidas), sua lógica BDI provavelmente (já que não foi provado) seria completa, mas não-correta.

O último ponto a ser explicitado é a divergência entre os autores com respeito à propriedade de Efeitos Indesejados. Rao e Georgeff demonstram que sua teoria formal soluciona este problema através da condição semântica, entretanto, sua teoria formal não é correta (conforme explicado anteriormente). Wooldridge, em [Woo00a], prova o seguinte teorema:

Teorema 4.4.1 *Sejam $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des ou $Goal$ e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R e $\overline{R}$, respectivamente. Se em um modelo, $R_t^w(i) \subseteq \overline{R}_t^w(i)$ então o esquema $\mathcal{M}_i(\varphi) \wedge \overline{\mathcal{M}}_i(\varphi \rightarrow \psi) \wedge \neg\mathcal{M}_i(\neg\psi)$ é insatisfazível neste modelo.*

A partir deste teorema, visualiza-se que o problema de Efeitos Indesejados não é solucionado quando adotada a propriedade do Realismo (qualquer relação de inclusão – $\subseteq$ – entre relações de acessibilidade entre mundos desse modelo). Incluindo, assim, o Realismo Forte definido por Rao e Georgeff nos artigos [RG91, RG93].

4.5 Soluções para as Lógicas BDI

Como foi visto, a causa das implicações 4.1 e 4.3 não expressarem sintaticamente as condições semânticas explicitadas aqui anteriormente, descritas por Rao e Georgeff e Wooldridge, é a inexistência de qualquer relação de inclusão nas definições das relações de acessibilidade e dos operadores modais. Portanto, apresentar-se-á duas possíveis soluções com o mesmo objetivo: relacionar os mundos de intenções (iw), de objetivos (gw) e de crenças (bw) de tal forma que $iw \sqsubseteq gw \sqsubseteq bw$.

Solução 1 Os operadores modais são definidos em [Woo00a, RG91, RG93] como:

$$M, V, w_t \models \overline{\mathcal{M}}(\varphi) \text{ sse } \forall w' \in W \text{ tal que } w_t R'_{\overline{\mathcal{M}}} w',\ M, V, w'_t \models \varphi$$

onde $\overline{\mathcal{M}}$ é o operador modal Bel, $Goal$ ou Int; e $R'_{\overline{\mathcal{M}}}$ a relação de acessibilidade correspondente.

Agora, propõe-se a seguinte interpretação:

$$\begin{aligned}
&M, V, w_t \models Bel(\varphi) \text{ sse } \forall w' \in W \text{ tal que } w_t B w',\ M, V, w'_t \models \varphi\\
&M, V, w_t \models Goal(\varphi) \text{ sse } \forall w' \in W \text{ tal que } w_t G w',\ M, V, w'_t \models \varphi\\
&\qquad\qquad \text{e } \exists w'' \in W \text{ tal que } w_t B w'' \wedge w' \sqsubseteq w''\\
&M, V, w_t \models Int(\varphi) \text{ sse } \forall w' \in W \text{ tal que } w_t I w',\ M, V, w'_t \models \varphi\\
&\qquad\qquad \text{e } \exists w'' \in W \text{ tal que } w_t G w'' \wedge w' \sqsubseteq w''
\end{aligned}$$

Note que os mundos de crenças não precisam ser sub-mundos do mundo do qual eles se originam, mesmo que este seja o suposto mundo "real". Em [RG91, RG93], Rao e Georgeff afirmam que mundos de crenças (w') devem corresponder ao mundo w do qual provém a relação B. Mas não especificam qual forma de correspondência eles devem ter. Na concepção dada por este trabalho, w' não deve necessariamente ser sub-mundo de w. É suficiente que exista em w' o estado de tempo t, tal que $w_t B w'$, com valores verdade das sentenças de acordo com a definição de Bel_i em w_t. Entretanto, pode ser interessante – dependendo do tipo de agente a ser construído (como por exemplo um robô) – que o agente acredite em tudo o que foi realizado no passado, assim, este agente armazenaria como crenças todas as suas experiências, ou seja, w' possuiria o mesmo passado (mesmos estados t_i e R que relacionam tais estados) de w. Contudo, manter-se-á o modelo de crenças mais genérico (definido como mostrado nas soluções 1 e 2).

Solução 2 Uma relação de acessibilidade R' é definida em [Woo00a] como

$$R' : \Sigma_{Ag} \rightarrow \wp\{W \times T \times W\}$$

Seja R' uma relação B ou G ou I.

Agora propõe-se a seguinte definição estrutural (seja L_{Gr} um conjunto finito ou enumerável composto por agentes e grupos de agentes):

$$B : L_{Gr} \rightarrow \wp\{W \times T \times W\}$$
$$G : L_{Gr} \rightarrow \wp\{W \times T \times W\}$$
$$I : L_{Gr} \rightarrow \wp\{W \times T \times W\}$$

Tais que satisfazem as seguintes condições:

- $B_t^w(i) \subseteq_{sup} G_t^w(i)$
- $G_t^w(i) \subseteq_{sup} I_t^w(i)$.

Apesar de ambas as soluções serem possíveis de serem aplicadas, a primeira é mais complexa, uma vez que ela propõe a criação de operadores modais definidos sobre outros operadores modais, ou seja, propõe-se a criação de uma lógica modal de alta ordem. Portanto, para simplificar, aplicar-se-á uma variação da solução 2 na nova lógica BDI a ser apresentada neste trabalho. A seguir ver-se-á a prova da bi-implicação entre as propriedades sintáticas e as semânticas relacionadas às atitudes mentais:

Lema 4.5.1 $\models Goal_i(optional(\varphi)) \rightarrow Bel_i(optional(\varphi))$.

PROVA: Pela definição de B e G apresentada na solução 2, tem-se que $B_t^w(i) \subseteq_{sup} G_t^w(i)$. Daí, se $w' \in B_t^w(i)$, então, existe um $w'' \in G_t^w(i)$ tal que $w'' \sqsubseteq w'$. Agora, assumindo que $\models Goal_i(optional(\varphi)) \rightarrow Bel_i(optional(\varphi))$ é falso, tem-se que para algum $\langle M, V, w, t\rangle$ arbitrário, $M, V, w_t \models Goal_i(optional(\varphi))$ e $M, V, w_t \models \neg Bel_i(optional(\varphi))$. Então, pela semântica de $Goal$ e Bel conclui-se, respectivamente, que para todo $x'' \in G_t^w(i)$, $M, V, w_t'' \models optional(\varphi))$ e existe um $w' \in B_t^w(i)$ tal que $M, V, w_t' \models \neg optional(\varphi)$. Entretanto, por consequência de $\overline{R}_t^w(i) \subseteq_{sup} R_t^w(i)$, se existe $w' \in B_t^w(i)$ tal que $M, V, w_t' \models \neg optional(\varphi)$, então, existe um $w'' \in G_t^w(i)$ tal que $M, V, w_t'' \models \neg optional(\varphi)$. O que é uma contradição. Logo, a suposição é falsa. ∎

Corolário 4.5.1 $B_t^w(i) \subseteq_{sup} G_t^w(i)$ *sse* $\models Goal_i(optional(\varphi)) \rightarrow Bel_i(optional(\varphi))$.

Lema 4.5.2 $\models Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$.

PROVA: Prova análoga à prova da proposição 4.5.1. ■

Corolário 4.5.2 $G_t^w(i) \subseteq_{sup} I_t^w(i)$ *sse* $\models Int_i(optional(\varphi)) \rightarrow Goal_i(optional(\varphi))$.

Em [Woo00a], há a afirmação de que as outras inter-relações entre $B_t^w(i)$, $G_t^w(i)$ e $I_t^w(i)$ (não apresentadas neste trabalho) também correspondem a outras propriedades sintáticas condicionais, envolvendo os modais Bel, $Goal$ e Int; todavia, pelo mesmo motivo apresentado nas proposições 4.4.1 e 4.4.2, as propriedades sintáticas condicionais entre Bel, $Goal$ e Int não geram a mesma interpretação das inter-relações semânticas entre $B_t^w(i)$, $G_t^w(i)$ e $I_t^w(i)$.

Os axiomas A3, A4 e A5 de Rao e Georgeff, apresentados em [RG93], requerem propriedades as quais não são representadas semanticamente, nem pela definição dos operadores modais ou suas respectivas relações de acessibilidade, nem pela interpretação desses operadores. Porém, elas podem ser substituídas pelo acréscimo da cláusula (‘e’) na definição de submundos apresentada no mesmo artigo. Assim, a definição de sub-mundo apresentada por Rao e Georgeff precisaria ser da seguinte forma:

Definição 4.5.1 *Seja $\overline{R'}$ uma relação de acessibilidade. Então, um mundo w' é submundo de w, denotado por $w' \sqsubseteq w$, se e somente se w' é uma sub-árvore (no significado lógico não usual da expressão sub-árvore) de w com as mesmas valorações para as sentenças, ou seja, sse (a) $T_{w'} \subseteq T_w$; (b) para todo $u \in T_{w'}$, $\Phi(q, w', u) = \Phi(q, w, u)$, onde q é um símbolo de predicado; (c) para todo $u \in T_{w'}$, $R_u^w = R_u^{w'}$; (d)$A_{w'}$ é igual a A_w quando este é restrito aos estados de $T_{w'}$ e similarmente para $S_{w'}$ e $F_{w'}$; e (e) $w\overline{R'}w'$.*

Assim, as condições semânticas dos axiomas A3, A4 e A5 apresentadas por Rao e Georgeff e explicitadas na sub-seção 4.2.3, seriam satisfeitas na semântica de sua lógica BDI pela definição 4.5.1 e pelas definições de B, G e I apresentada na solução 2.

O item (e) da definição 4.5.1 a torna muito forte. E para tanto, ela requer um estudo ontológico a fim de verificar a sua real necessidade. Contudo, a definição explicitada na solução 2 não será a exata solução aplicada na lógica BDI desenvolvida neste trabalho. Tal definição é apresentada, apenas, para solucionar as ambiguidades existentes nas lógicas BDI desenvolvidas por Rao e Georgeff. Portanto, não preocupar-se-á em apresentar justificativas ontológicas (se é que existem) para a adição do item ’e’ à definição original de sub-mundo apresentada por Rao e Georgeff.

4.6 Considerações Finais

No capítulo anterior, foram apresentadas formalizações lógicas que podem ser usadas na representação de arquiteturas de agentes cognitivos. Uma das formalizações (formalização de Cohen e Levesque), em particular, utiliza crenças, desejos e *p-goal* (objetivos persistentes que podem ser pensados como intenções, caso intenção seja considerada uma forma de compromisso sobre objetivos) também é uma lógica BDI. Todavia, apenas as lógicas BDI apresentadas neste capítulo utilizam especificamente as atitudes mentais crença, desejo/objetivo e intenção e são baseadas unicamente na teoria filosófica intencional dirigida ao raciocínio prático de Bratman. Com respeito a estas lógicas BDI, foram apontadas algumas divergências entre a semântica e a sintaxe de ambas de modo que se conclui as incorretudes do sistema formal BDI de Rao e Georgeff e de LORA (caso fosse construído um sistema formal para tal lógica) quaisquer que sejam as propriedades sintáticas escolhidas para definir as relações entre atitudes mentais. Apresentou-se, ainda, neste capítulo, duas possíveis soluções para tais problemas e provou-se que os mesmos são resolvidos através de uma dessas soluções (solução 2) e da nova definição de sub-mundos.

Entretanto, tal como é lembrado por Wooldridge em [Woo00a, 100–101,109], o realismo forte é uma propriedade inaceitavelmente muito forte para a maioria dos agentes. Esta opinião é compartilhada por Bratman o qual apresenta uma tese de assimetria para melhor tratar as inter-relações entre crenças e intenções. Por isso, uma nova propriedade (um realismo enfraquecido) deve ser definido para expressar sintaticamente a inter-relação entre objetivos e crenças. Wooldridge argumenta que esta inter-relação seria mais aceitavelmente expressa por $Goal_i(optional(\varphi)) \rightarrow \neg Bel_i(\neg optional(\varphi))$ o que corresponderia semanticamente à $B_t^w(i) \cap_{sup} G_t^w(i) \neq \emptyset$ (e vice-versa). Considerando esta inter-relação entre objetivos e crenças, estabelece-se uma nova relação entre objetivos e intenções, o que acarretará em uma nova relação entre intenções e crenças concordante à tese de assimetria.

É importante ressaltar que para fazer a análise das propriedade sintáticas e semânticas entre as atitudes mentais, apresentou-se neste capítulo as definições de um conjunto dessas propriedades. Algumas dessas propriedades não haviam sido definidas por Woldridge [Woo00a] nem por Rao e Georgeff, bem como, havia divergência, em cada trabalho, entre as definições dos dois autores e dentro do próprio trabalho – como é o caso de $R_t^w(i) \subseteq_{sup} \overline{R}_t^w(i)$ definido por Wooldridge em [Woo00a] de duas formas distintas.

No próximo capítulo, mostrar-se-á que essa inter-relação – referenciada por Wooldridge como a mais aceitável – é a formalização a qual deve ser adotada por uma lógica BDI fundamentada segundo a tese de assimetria de Bratman. Provar-se-á, ainda, no mesmo capítulo, a correspondência entre as novas relações semânticas entre B, G e I e as novas propriedades sintáticas entre Bel, $Goal$ e Int – os quais são redefinidos para que as atitudes possuam as propriedades requeridas pela tese de assimetria.

Capítulo 5

Uma Nova Lógica BDI

As inter-relações entre as atitudes mentais são os pontos mais relevantes para definir uma teoria de intenções para agentes inteligentes. Bratman construiu uma teoria de intenções cujas inter-relações entre crenças, desejos e intenções são baseadas na definição de cada uma delas (como a psicologia popular entende ou define crenças, desejos e intenções), no modelo de crença-desejo e na tese de assimetria. Os dois primeiros tópicos já foram abordados ao longo do trabalho. Tratar-se-á agora da tese de assimetria para que, em seguida, seja definida uma lógica BDI (abordagens sintática e semântica) concordante com os princípios descritos por sua fundamentação filosófica ([Bra87]) e tentando garantir também a corretude e a completude da mesma.

5.1 Tese de Assimetria de Bratman

Como aspecto fundamental da semântica, em [RG91, RG93], há a suposição da propriedade de que, para que a intenção de um agente seja consistente com a crença do mesmo, é necessário assegurar que uma intenção de A requer uma crença em A – propriedade esta forçada semanticamente pela relação de submundos entre os mundos de intenções com relação aos mundos de crença (4.2 quando utilizada a solução 2) e forçada sintaticamente pelo axioma A1 – todavia, Bratman não quer usar uma suposição tão forte e apresenta razões para não aceitá-la. Ele em [Bra87, 37–41], defende a tese de assimetria apresentando dois exemplos relevantes:

Sou agnóstico a respeito da intenção Eu pretendo agora parar na livraria, quando eu estiver de bicicleta a caminho de casa. Mesmo sabendo da tendência de eu não parar, uma vez que quando eu ando de bicicleta eu ligo o "piloto-automático" até chegar em casa e por isso eu posso me esquecer de parar na livraria. Note que isso não é acreditar que eu não vou parar, eu apenas não acredito que vou.

Agnóstico a respeito dos sucessos Talvez eu pretenda realizar uma operação de alto risco a qual requer uma série de etapas difíceis. Estou confiante que darei o meu melhor, mas tenho dúvidas sobre o sucesso dessas etapas. Note que eu não acredito na falha da tentativa, mas também não acredito no sucesso.

Avaliando tais exemplos, percebe-se que pretender A não requer necessariamente uma crença de que alguém fará A, ou seja, ter a intenção de α não requer acreditar que α será realizado. Segundo Bratman, pretender A normalmente fornece ao agente suporte à crença de que ele fará A, todavia, há alguns casos em que isto não ocorre, isto é: normalmente

$$\models IntA \rightarrow BelA \quad (5.1)$$

Mas, não há irracionalidade alguma (conforme mostrado nos exemplos) caso um agente pretenda A e não acredite que fará A (por esquecimento ou simplesmente por duvidar do sucesso de realizar A) – o que é chamado por Bratman de incompletude de intenção-crença –, ou seja:

$$\models Int\,A \wedge \neg Bel\,A \quad (5.2)$$

Mas, seria irracional um agente pretender A e acreditar que não fará A – Bratman chama isto de irracionalidade de intenção-crença –, ou seja:

$$\not\models Int\,A \wedge Bel\,\neg A \quad (5.3)$$

Destaca-se, então, duas propriedades: $IntA \wedge \neg BelA$ (incompletude de intenção-crença) e $\neg(Int\,A \wedge Bel\,\neg A) \equiv IntA \rightarrow \neg Bel \neg A$ (irracionalidade de intenção-crença) que, por sua vez, implica em 5.3. Essas duas propriedades são inferidas respectivamente por $B_t^w \subseteq_{sup} I_t^w$ (conforme a proposição 4.3.3) e por $I_t^w(i) \cap_{sub} B_t^w(i) \neq \emptyset$ (conforme a proposição 4.3.4) – sendo A uma O-fórmula.

Portanto, vê-se que há sentenças as quais pertencem aos mundos de intenções e de crenças em função dos primeiros serem sub-mundos do segundo; mas pode haver também mundos de intenções que não são submundos dos mundos de crença, ou seja, $\forall w''.\exists w'.(w' \in I_t^w \wedge w'' \in B_t^w \rightarrow w' \not\sqsubseteq w''$ o que é inferido por $Int\,A \wedge \neg Bel\,A$. Enquanto não é possível que haja uma sentença num mundo de intenções cuja contradição esteja num mundo de crenças (equação 5.3 a qual é proveniente de $\vDash \neg(Int\,A \wedge Bel\,\neg A)$ o que é equivalente a $\vDash Int\,A \rightarrow \neg(Bel\,\neg A)$). A partir de tais premissas e devido a forma com que os mundos, sub-mundos e modelos são definidos nas lógicas BDI, infere-se que algum(ns) mundo(s) de intenções é(são) sub-mundo(s) do(s) mundo(s) de crenças e (possivelmente) outros não. Ou seja, a inter-relação entre as intenções e as crenças há de ser uma intersecção. Para enfraquecer tal intersecção, define-se a condição semântica $I_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$ definida em [Woo00a] como: $I_t^w(\imath) \cap_{sub} B_t^w(\imath) = \{w' \mid w' \in B_t^w(\imath)$ e $\exists w'' \in I_t^w(\imath)$ t.q. $w'' \sqsubseteq w'\}$. O que implica em $Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Enfim, a lógica BDI definida neste capítulo tem a pretensão de ser uma formalização da teoria de Bratman e, por isso, deve satisfazer os dois princípios da tese de assimetria os quais, como foi visto no parágrafo anterior, são obtidos através da relação de intersecção não vazia entre os mundos de crenças e os mundos de intenções.

Bordini e Moreira, em [BM02], também concluem esses dois princípios – $\models Int\varphi \rightarrow \neg(Bel\neg\varphi)$ e $\not\models Int\varphi \rightarrow Bel\varphi$ – a partir da tese de assimetria; e afirmam que Rao e Georgeff, em [RG98], acrescentam mais uma propriedade – $\not\models BelA \rightarrow IntA$ – para construir todas as possíveis relações entre crenças e intenções. Esta, por sua vez, é, na verdade, uma consequência do princípio descrito na equação 5.2. Posteriormente, Bordini e Moreira afirmam que Rao e Georgeff, no mesmo artigo, acrescentam seis propriedades (princípios) semelhantes para construir as relações entre crenças e desejos e entre desejos e intenções. Ou seja, Rao e Georgeff, em [RG98], definem nove princípios para a Tese de Assimetria de Bratman, são eles:[1]

- $\models Int\varphi \rightarrow \neg(Bel\neg\varphi)$
- $\not\models Int\varphi \rightarrow Bel\varphi$
- $\not\models Bel\varphi \rightarrow Int\varphi$
- $\models Int\varphi \rightarrow \neg(Goal\neg\varphi)$
- $\not\models Int\varphi \rightarrow Goal\varphi$
- $\not\models Goal\varphi \rightarrow Int\varphi$
- $\models Goal\varphi \rightarrow \neg(Bel\neg\varphi)$
- $\not\models Goal\varphi \rightarrow Bel\varphi$
- $\not\models Bel\varphi \rightarrow Goal\varphi$

A relação entre os objetivos e as intenções requerida na lógica BDI proposta aqui é a mesma definida nas lógicas de Rao e Georgeff apresentadas em [RG91, RG93]. Isto é, as intenções são objetivos os quais o(s) agente(s) se compromete(m) em satisfazer. Assim, se φ é uma intenção então φ também é um objetivo. E para uma melhor adequação à semântica de mundos possíveis com futuro ramificado, diz-se que pretender opcionalmente φ tem por conseguinte objetivar opcionalmente φ. Sugerindo, portanto, uma relação de inclusão $\subseteq_{sub}$ entre I e G – $I_t^w \subseteq_{sub} G_t^w$. Para forçar essas relações entre as atitudes mentais, define-se as relações de acessibilidade com as seguintes condições de satisfação.

Definição 5.1.1 $\begin{array}{l} B : L_{Gr} \rightarrow \wp\{W \times T \times W\} \\ G : L_{Gr} \rightarrow \wp\{W \times T \times W\} \\ I : L_{Gr} \rightarrow \wp\{W \times T \times W\} \end{array}$ *Tais que satisfazem as seguintes condições:*

- $\exists w', w''.(w' \in B_t^w(\imath) \wedge w'' \in G_t^w(\imath) \wedge w'' \sqsubseteq w')$*;*
- $\forall w'.(w' \in G_t^w(\imath) \rightarrow \exists w''.(w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq w')$*; e*
- $\exists w', w''.(w' \in B_t^w(\imath) \wedge w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq w')$.

Na seção a seguir, apresentar-se-á a proposta de uma nova lógica BDI a qual utiliza a definição 5.1.1 para definir as relações de acessibilidade B, G e I. Por consequência, tem-se que esta nova lógica captura a noção da tese de assimetria apresentada por Bratman. E a prova desta afirmação será vista ao longo da seção.

[1] O operador usado no artigo de Bordini e Moreira era o Des ao invés do $Goal$, mas como se está usando a modalidade $Goal$ para a lógica BDI clássica, os princípios são aqui apresentados com o operador $Goal$.

5.2 Proposta de uma Nova Lógica BDI

Nesta seção apresentar-se-á a proposta de uma nova lógica BDI baseada nas duas outras versões da lógica BDI, a de Rao e Georgeff e a de Wooldridge (LORA), mostradas no capítulo anterior sem os problemas explicitados anteriormente. Para solucionar tais problemas, usar-se-á uma variante da solução 2, ou seja, a definição das relações de acessibilidade segundo a equação 5.1.1.

5.2.1 Sintaxe da Lógica BDI

A linguagem da lógica BDI é definida em três fases, a primeira define uma linguagem para eventos $\mathcal{L}^E = \langle \Sigma^E, \mathcal{G}^E \rangle$, onde $\Sigma^E = \Sigma_{Ac} \cup \Sigma_{L^E}$ e $\mathcal{G}^E = \{E_1, E_2, E_3, E_4\}$ tal que

$\Sigma_{Ac} = \{\alpha, \beta, \gamma, \alpha_1, \alpha_2, \alpha_3 \ldots, \beta_1, \beta_2, \beta_3, \ldots, \gamma_1, \gamma_2, \gamma_3, \ldots\}$ é um conjunto enumerável de ações e

$\Sigma_{L^E} = \{\ ;\ ,\ |,\ *\}$ é um conjunto de símbolos aplicados sobre ações.

E_1 $\dfrac{-}{\alpha}$

E_2 $\dfrac{\alpha,\ \beta}{\alpha;\beta}$

E_3 $\dfrac{\alpha,\ \beta}{\alpha|\beta}$

E_4 $\dfrac{\alpha}{\alpha*}$

A segunda fase define uma linguagem de grupos de agentes $\mathcal{L}^{Gr}$ tal que $\mathcal{L}^{Gr} = \langle \Sigma_G \cup \{\in\} \cup \Sigma_{Ag}, \mathcal{G}^{Gr} \rangle$, onde

$\Sigma_{Ag} = \{i_1, i_2, i_3 \ldots\}$ é um conjunto enumerável de agentes

$\Sigma_G = \{gr, gr_i \,|\, i \geq 1\}$ é um conjunto enumerável de grupos de agentes

e $\mathcal{G}^{Gr} = \{G_1, G_2\}$, onde

G_1 $\dfrac{-}{i}$, onde $i \in \Sigma_{Ag}$

G_2 $\dfrac{i_1, \ldots, i_n,\ gr}{i_1, \ldots, i_n \in gr}$, onde $i_1, \ldots, i_n \in \Sigma_{Ag}$ e $gr \in \Sigma_G$.

E a terceira fase define a linguagem da lógica BDI $\mathcal{L}^{3M}$. Ela é constituída pelo alfabeto $\Sigma^{3M} \cup Ling(\mathcal{L}^E) \cup Ling(\mathcal{L}^{Gr}), \mathcal{G}^{3M}$ onde

$\Sigma^{3M} = \langle X \cup \Sigma_d \cup \Sigma_P \cup \Sigma_{L^{3M}} \rangle$;

L^E é um conjunto enumerável de fbf de $\mathcal{L}^E$, ou seja, é um conjunto enumerável de eventos primitivos ou não;

$\Sigma_{L^{3M}} = \Sigma_{L'} \cup \{\bigcirc, \mathcal{U}, \mathcal{W}, ?, Inev, Opt, Bel, Goal, Int, Success, Fail, Execute, Group\}$ é o alfabeto de símbolos lógicos;[2]

$\mathcal{G}^{3M} = F_1, \ldots, F_{26}$, onde:

$F_1, \ldots, F_8$ Definidos na seção 3.2.1.

F_9, F_{10} Definidos na seção 3.3.

F_{11} $\dfrac{\varphi}{\bigcirc\varphi}$

F_{12} $\dfrac{\varphi,\ \phi}{\varphi\mathcal{U}\phi}$

F_{13} $\dfrac{\varphi,\ \phi}{\varphi\mathcal{W}\phi}$

F_{14} $\dfrac{\varphi,\ \imath}{Bel_\imath(\varphi)}$

F_{15} $\dfrac{\varphi,\ \imath}{Goal_\imath(\varphi)}$

F_{16} $\dfrac{\varphi,\ \imath}{Int_\imath(\varphi)}$

F_{17} $\dfrac{\varphi}{Inev(\varphi)}$

F_{18} $\dfrac{\varphi}{Opt(\varphi)}$

F_{19} $\dfrac{e_1,\ e_2}{e_1; e_2}$

F_{20} $\dfrac{e_1,\ e_2}{e_1 | e_2}$

F_{21} $\dfrac{e_1}{e_1*}$

F_{22} $\dfrac{\varphi}{\varphi?}$

F_{23} $\dfrac{\varphi,\ e_1,\ \imath}{Success_{i_1,\ldots,i_n}(e_1,[\varphi])}$

F_{24} $\dfrac{\varphi,\ e_1,\ \imath}{Fail_{i_1,\ldots,i_n}(e_1,[\varphi])}$

F_{25} $\dfrac{\varphi,\ e_1,\ \imath}{Execute_{i_1,\ldots,i_n}(e_1,[\varphi])}$

F_{26} $\dfrac{\varphi,\ e_1,\ gr}{Group(gr)(e_1,[\varphi])}$

Observação 5.2.1 *1. O conjunto L^{3M} é o conjunto de fórmulas bem formadas de $\mathcal{L}^{3M}$. Assim $L^{3M} = Ling(\mathcal{L}^{3M})$*

2. *$\imath = \{i_1, \ldots i_n\}$, onde $n \geq 1$.*

3. *Nota-se que seria possível diminuir tal linguagem em função da definição de operadores derivados. Entretanto, é desejável manter o conjunto $\Sigma_{L^{3M}}$, pelo menos na abordagem semântica, para evidenciar as interpretações desses conectivos lógicos.*

4. *Nesta lógica também se prefere a utilização de $Goal$ à utilização de Des uma vez que se desenvolve aqui uma lógica cujos princípios do terceiro-excluído e da [não-]contradição são obedecidos, logo, não é desejável trabalhar com desejos conflitantes ou contraditórios gerando intenções contraditórias e tornando, portanto, o sistema inconsistente.*

[2] $\Sigma_{L'}$ é definido na seção 3.3.

5.2.2 Semântica da Nova Lógica BDI

A semântica desta lógica é uma semântica de Kripke dada pelo modelo M o qual é definido pela tupla $\langle W, T, R, B, G, I, \mathbf{D}, \Phi, E\rangle$ cujos componentes são descritos semelhantemente aos elementos de M na sub-seção 4.2.2 (sendo D definido como $\mathbf{D} = \langle D_{Ac}, D_{Ag}, D_{Gr}, D_U\rangle$ denotando, assim, o domínio de discurso onde D_U é um domínio de outros objetos não pertencentes aos domínios anteriores). Há, ainda, a adição do conjunto de eventos E na tupla o qual não tinha lugar nas outras formalizações.

Entretanto, o modelo de tempo desta lógica BDI, as definições de mundo e sub-mundo são similar ao definido por Rao e Georgeff em [RG91, RG93]. Daí tem-se para esta lógica BDI as seguintes definições:

1. A semântica de Kripke é um modelo $M = \langle W, T, R, B, G, I, \mathbf{D}, \Phi, E\rangle$. Onde W, T, $\mathbf{D}$ e Φ são definidos similarmente a [RG91, RG93] (explicitado na sub-seção 4.2.2); R é um conjunto de relações (arcos) entre esses estados, onde essas relações são totais, transitivas, lineares no passado e ramificadas no futuro $\mathbf{D} = \langle D_{Ac}, D_{Ag}, D_{Gr}, D_U\rangle$ é um domínio de discurso composto por ações, agentes, grupos de agentes e outros objetos; e B, G ou I são relações – de crença, objetivo e intenção, respectivamente – entre mundos ($B \subseteq W \times W$, $G \subseteq W \times W$ e $I \subseteq W \times W$). E é um conjunto de eventos 'e' os quais podem ser primitivos ou não.

2. Um mundo w é uma tupla $w = \langle T_w, R_w, Suc_w, Fal_w\rangle$ onde $T_w \subseteq T$ é um conjunto de estados de tempo de w, $R_w \subseteq R$ é um conjunto de arcos que ligam estados de tempo adjacentes de w. Suc_w e Fal_w são respectivamente funções cuja definição é semelhante à definição de S_w e F_w em [RG91, RG93] (também explicitado na sub-seção 4.2.2), no entanto, $Suc_w : D_{Ag} \times T_w \times T_w \rightarrow E$ está relacionada a um dado agente ou grupo de agentes; e (conforme dito em tais artigos) possuem as propriedades de simetria, ou seja: $Suc_w(\imath)(t_i, t_j) = Suc_w(\imath)(t_j, t_i)$ e $Fal_w(\imath)(t_i, t_j) = Fal_w(\imath)(t_j, t_i)$.

3. Um mundo w' é submundo de w, denotado por $w' \sqsubseteq w$, se e somente se: (a) $T_{w'} \subseteq T_w$; (b) para todo $u \in T_{w'}$, $\Phi(q, w', u) = \Phi(q, w, u)$, onde q é um símbolo de predicado; (c) para todo $u \in T_{w'}$, $R_u^w = R_u^{w'}$; (d) $A_{w'}$ é igual a A_w quando este é restrito aos estados de $T_{w'}$ e similarmente para $Suc_{w'}$ e $Fal_{w'}$.

Definição 5.2.1 *Seja a situação (w, t) em um dado modelo; caso haja um mundo de intenções $w''' \in I_t^w$ então haverá mundo(s) de crenças $w' \in B_t^w$ e mundo(s) de objetivos $w'' \in G_t^w$ tais que $w'Iw'''$ e $w''Iw'''$. E havendo um mundo de objetivos $w'' \in G_t^w$, há mundo(s) de crenças $w'B_t^w$ tal que $w'Gw''$. As relações de acessibilidade B, G, e I, são definidas conforme exposto em 5.1.1. Os elementos de G_t^w se relacionam aos mundos de I_t^w via relação I e os elementos de B_t^w se relacionam aos mundos de G_t^w e de I_t^w via relação G e I, respectivamente. Isto é, há uma injeção entre os conjuntos G_t^w e I_t^w, entre B_t^w e G_t^w e entre B_t^w e I_t^w.*

A partir dessa definição, será possível construir uma lógica BDI baseada nos modelos de crença-desejo e tese de assimetria definidos na teoria de Bratman. Bem como, possibilitará o agente ou grupo de agentes acreditar que tem uma intenção e um dado desejo assim que ele tiver tal intenção ou desejo; e desejar sua intenção, quando tiver alguma – propriedade requerida sintaticamente pelos axiomas A3, A4 e A5 explicitados anteriormente.

Definição 5.2.2 *Para cada mundo w e atribuição de valores às variáveis $\rho^w : X \rightarrow D_w$ temos uma valoração dos termos da linguagem dada pela função $V_{\rho^w} : Ling(\mathcal{L}^T) \rightarrow D_w$ definida por:*

$$V_{\rho^w}(x) = \rho^w(x) \text{ para cada } x \in X$$

$$V_{\rho^w}(a) = a_w \text{ para cada } a \in \Sigma_C$$

$$V_{\rho^w}(f(\tau_1, \ldots \tau_n)) = f_w(V_{\rho^w}(\tau_1), \ldots, V_{\rho^w}(\tau_n))$$

onde a_w e f_w são as interpretações dos símbolos constante a e de função f no mundo w, respectivamente.

Seja V é uma função de valoração que mapeia uma situação de um dado agente ao conjunto $\{0, 1\}$, ou mais formalmente: $V : S_w \times L^{3M} \rightarrow \{0, 1\}$, onde $S_w = \{(w, t) \mid w \in W \text{ e } t \in T_w\}$.

$$\begin{aligned} &V_t^w(R(\tau_1, \ldots \tau_n)) = R_w(V_{\rho^w}(\tau_1), \ldots, V_{\rho^w}(\tau_n)) \text{ para cada } R \in \Sigma_R \\ &\text{com } arid(R) = n \text{ e } \tau_1 \ldots \tau_n \in Ling(\mathcal{L}^{3M}) \end{aligned}$$

onde R_w é a interpretação do símbolo de predicado R no mundo w.

$$V_t^w(\neg\varphi) = \begin{cases} 1 & \text{, se } V_t^w(\varphi) = 0 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\varphi \vee \phi) = \begin{cases} 1 & \text{, se } V_t^w(\varphi) = 1 \text{ ou } V_t^w(\phi) = 1 \text{ ou ambos} \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\varphi \wedge \phi) = \begin{cases} 1 & \text{, se } V_t^w(\varphi) = 1 \text{ e } V_t^w(\phi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\varphi \rightarrow \phi) = \begin{cases} 0 & \text{, se } V_t^w(\varphi) = 1 \text{ e } V_t^w(\phi) = 0 \\ 1 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\varphi \leftrightarrow \phi) = \begin{cases} 1 & \text{, se } V_t^w(\varphi) = V_t^w(\phi) \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\forall x.\phi) = \begin{cases} 1 & \text{, se para toda atribuição } V_t'^w : X \rightarrow U \text{ tal que} \\ & V_t'^w(y) = V_w(y) \forall y.(y \neq x)\, V_t'^w(\varphi) = 1 \\ 0 & \text{, caso contrário} \end{cases}$$

$$V_t^w(\exists x.\phi) = \begin{cases} 1 & \textit{, se para alguma atribuição } V_t'^w : X \rightarrow U \textit{ tal que} \\ & V_t'^w(y) = V_t^w(y)\, \forall y.(y \neq x)\, V_t'^w(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(\Box\phi) = \begin{cases} 1 & \textit{, se para todo } t' \mid tRt',\ V_{t'}^w(\phi) = 1 \\ 0 & \textit{, se para algum } t' \mid tRt',\ V_{t'}^w(\phi) = 0 \end{cases}$$

$$V_t^w(\Diamond\phi) = \begin{cases} 1 & \textit{, se para algum } t' \mid tRt',\ V_{t'}^w(\phi) = 1 \\ 0 & \textit{, se para todo } t' \mid tRt',\ V_{t'}^w(\phi) = 0 \end{cases}$$

$$V_t^w(\bigcirc\phi) = \begin{cases} 1 & \textit{, se para todo } c(1) \mid tRc(1),\ V_{c(1)}^w(\phi) = 1 \\ 0 & \textit{, se para algum } c(1) \mid tRc(1),\ V_{c(1)}^w(\phi) = 0 \end{cases}$$

$$V_t^w(\varphi\,\mathcal{U}\,\phi) = \begin{cases} 1 & \textit{, se existir um } u \in c, \textit{ tal que } V_{c(u)}^w(\phi) = 1, V_{c(u)}^w(\varphi) = 0 \\ & \textit{e para todo } 0 \leq t < u,\ V_t^w(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(\varphi\mathcal{W}\phi) = \begin{cases} 1 & \textit{, se existir um u tal que } u \in c,\ V_{c(u)}^w(\phi) = 1, V_{c(u)}^w(\varphi) = 0 \textit{ e para todo} \\ & 0 \leq t < u,\ V_t^w(\varphi) = 1 ou\ se\ se\ \neg\exists u \textit{ então } \forall t \in c,\ V_t^w(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Bel_\imath(\varphi)) = \begin{cases} 1 & \textit{, se todo } w' \textit{ tal que } wBw',\ V_t^{w'}(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Goal_\imath(\varphi)) = \begin{cases} 1 & \textit{, se todo } w' \textit{ tal que } wGw',\ V_t^{w'}(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Int_\imath(\varphi)) = \begin{cases} 1 & \textit{, se todo } w' \textit{ tal que } wIw',\ V_t^{iw}(\varphi) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Inev(\varphi)) = \begin{cases} 1 & \textit{, se para todos os caminhos c de } w\ V_{c(t')}^w(\varphi) = 1 \\ & \textit{onde } t, t' \in T_w \textit{ e } t' > t \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Opt(\varphi)) = \begin{cases} 1 & \textit{, se para algum caminho c de } w,\ V_{c(t')}^w(\varphi) = 1 \\ & \textit{onde } t, t' \in T_w \textit{ e } t' > t \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V^w_{tn}(Sucess_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in c \textit{ tal que } i < c \textit{ e } Suc_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V^w_{tn}(Fail_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in c \textit{ tal que } i < c \textit{ e } Fal_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V^w_{tn}(Execute_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in c \textit{ tal que } i < c \textit{ e } Suc_w(\imath, ti, tn) = e \\ & \quad \textit{ou } Fal_w(ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V^w_t(Group(gr)(e)) = \begin{cases} 1 & \textit{, se existe } tn \in c \textit{ tal que } tn > t \textit{ e } V^w_{tn}Execute_{gr}(e) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

Observação 5.2.2

Onde $c(1)$ é o primeiro instante de tempo do caminho c e, $c(t)$ e $c(u)$ são, respectivamente, o t-ésimo e o u-ésimo estado do caminho c.

Há dois tipos de fórmulas nesta linguagem tal como é definido em [RG91, RG93]): fórmulas de estado e fórmulas de caminho. Todavia, há apenas uma função de valoração de fórmulas a qual valora tais fórmulas a partir de uma dada situação (w, t) independente delas serem de estado ou de caminho. Logo, as fórmulas de caminho são valoradas a partir dos valores da mesma em cada (w, t), isto é,

$$V^w_c(\varphi) = \bigwedge\{V^w_{ti}(\varphi) \,|\, ti \in c\}.$$

A partir de então, é possível determinar a satisfatibilidade ou não das fórmulas BDI.

Definição 5.2.3 *Seja um modelo M, uma função de valoração V e uma fórmula $\varphi \in L^{3M}$. φ é satisfazível para um modelo M (ou M satisfaz φ) se existe uma situação (w, t) do modelo M tal que $t \in T_w$, $w \in W$ e $V^w_t(\varphi) = 1$, denotado por, $M, w_t \models^{3M} \varphi$; e φ é insatisfazível em M, denotado por $M \not\models^{3M} \varphi$, se não existe uma situação deste modelo M tal que $V^w_t(\varphi) = 1$, ou em outras palavras: $V^w_t(\varphi) = 0$ para qualquer elemento de S_w. φ é verdadeira em um modelo M, denotado por $M \models^{3M} \varphi$ se existe pelo menos uma situação (w, t) em cada mundo possível w de M tal que $V^w_t(\varphi) = 1$, ou seja, φ é satisfeita em todos os mundos w de M; φ é contingente caso haja pelo mundos w e w' de M tais que, para qualquer instante t, $V^w_t(\varphi) = 1$ e $V^{w'}_t(\varphi) = 0$, isto é, φ é contingente sse $M, w, t \models^{3M} \varphi$ e $M, w' \not\models^{3M} \varphi$; e φ é falsa para o modelo M, denotado por $M, w \not\models^{3M} \varphi$, se para todos os elementos de S_w, $V^w_t(\varphi) = 0$. Por fim, φ é válida, denotado por $\models^{3M} \varphi$ sse para qualquer instante, mundo e modelo $V(\varphi) = 1$ – o conjunto das fórmulas válidas de $\mathcal{L}^{3M}$ é definido por $FV3M = \{\varphi \in L^{3M} / \models^{3M} \varphi\}$ –; e é insatisfazível quando não existe um modelo para a mesma, isto é, para qualquer $\langle M, w, t\rangle$, φ é falsa, denotado por $\not\models^{3M} \varphi$.*

A validade também pode ser dada em um dado conjunto de modelos (em uma dada classe C), denotado por $C \models^{3M} \varphi$.

Da mesma forma que há a definição de uma fórmula satisfazível, define-se a satisfazibilidade de um conjunto de fórmulas Γ de L^{3M} como sendo um modelo e uma situação (w, t) à qual todas as fórmulas $\varphi \in \Gamma$ são verdadeiras, ou seja, $V_t^w(\varphi) = 1$. Dada as definições anteriores, pode-se definir o conceito de consequência semântica e lógica BDI.

Definição 5.2.4 *Seja a linguagem modal $\mathcal{L}^{3M}$ e $\Gamma \subseteq L^{3M}$. φ é uma consequência semântica de Γ, denotado por $\Gamma \models^{3M} \varphi$ se todas as situações em que Γ for satisfazível φ também é, ou seja, $\forall (w,t)$ / se $\forall \beta \in \Gamma$, $M, w, t \models \beta$ então $M, w, t \models \varphi$.*

Definição 5.2.5 *Seja a linguagem modal $\mathcal{L}^{3M}$, a lógica* BDI*, denotada por Log^{3M} é a estrutura munida de sua linguagem modal e sua consequência semântica:*

$$Log^{3M} = \langle L^{3M}, \models^{3M} \rangle$$

A nova Lógica BDI como uma Formalização da Tese de Assimetria: Ver-se-á, agora, que as propriedades sintática e semântica acerca das atitudes mentais requeridas pela teoria de Bratman – modelo desejo-crença e tese de assimetria – são satisfeitos.

Proposição 5.2.1 $\models Goal_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Pela definição de G e B em 5.1.1, tem-se que $G_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$, então, existem $w'' \in B_t^w(\imath)$ e $w' \in G_t^w(i)$ tal que $w' \sqsubseteq w''$. Agora, assuma que $\models Goal_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Portanto, para algum $\langle M, V, w_t \rangle$ arbitrário,$M, V, w_t \models Goal_\imath(\varphi)$ e $M, V, w_t \models Bel_\imath(\neg Opt(\varphi))$. A partir da interpretação do conectivo modal $Goal$, conclui-se que para todo $x' \in G_t^w(\imath)$, $M, V, x'_t \models \varphi$ e pela semântica de Bel, sabe-se que para todo $y'' \in B_t^w(\imath)$, $M, V, y''_t \models \neg Opt(\varphi)$ e por consequência de $G_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$ existe pelo menos um $w' \in G_t^w(\imath)$ tal que $M, V, w'_t \models \neg Opt(\varphi)$ ou existe um $w'' \in B_t^w(\imath)$ tal que $M, V, w''_t \models Opt(\varphi)$. Portanto, há uma contradição. Logo, a suposição há de ser falsa. ∎

Corolário 5.2.1 $G_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$ *sse* $Goal_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

A inter-relação entre o conjunto de mundos de objetivos e o conjunto de mundos de intenções é a mesma explicitada por Rao e Georgeff nos artigos [RG91, RG93]. Pois, nesses artigos, a intenção é um objetivo o qual se tem o compromisso de cumprir; similar ao que é apresentado por Bratman:

Objetivo: é uma atitude influenciadora de conduta (influencia o agente, ou seja, dá opções de ações ao agente);

Intenção: é uma atitude controladora de conduta (determina como o agente vai agir – ou, simplesmente, o que ele vai realizar). Em outras palavras, ela é um desejo com um certo compromisso volicional (manutenção da intenção) e dirigido aos meios (devem ser definidas outras intenções para que se consiga realizar a intenção prévia).

O que torna a definição de Rao e Georgeff perfeitamente cabível, já que estes definem objetivos como um conjunto consistente de desejos e intenções como objetivos aos quais o agente se compromete em realizar sem haver possibilidade de redução de um no outro (definir um em função do outro como feito por Cohen e Levesque com o conectivo $P - Goal$). Mostrar-se-á a seguir a correspondência entre as relações sintática e semântica – entre objetivo e intenção – requeridas pela teoria de Bratman.

Proposição 5.2.2 $\models Int_\imath(Opt(\varphi)) \rightarrow Goal_\imath(Opt(\varphi))$.

PROVA: Pela definição de I e G em 5.1.1, tem-se que $G^w_t(\imath) \subset_{sup} I^w_t(\imath)$, então, para todo $x' \in G^w_t(\imath)$ existe algum $w'' \in I^w_t(\imath)$ onde $w'' \sqsubseteq x'$. Agora assuma que não é o caso de $\models Int_\imath(Opt(\varphi)) \rightarrow Goal_\imath(Opt(\varphi))$; portanto, para um $\langle M, V, w, t\rangle$ qualquer, $M, V, w_t \models Int_\imath(Opt(\varphi))$ e $M, V, w_t \models \neg Goal_\imath(Opt(\varphi))$. A partir da interpretação dos conectivos Int e $Goal$, sabe-se, respectivamente, que: para todo $y'' \in I^w_t(\imath)$, $M, V, y''_t \models Opt(\varphi)$ e existe um $w' \in G^w_t(\imath)$ tal que $M, V, w'_t \models \neg Opt(\varphi)$. Mas como $G^w_t(\imath) \subset_{sup} I^w_t(\imath)$ e existe um $w' \in G^w_t(\imath)$, $M, V, w'_t \models \neg Opt(\varphi)$, segue que existe um $w'' \in I^w_t(\imath)$, $M, V, w''_t \models \neg Opt(\varphi)$. Isto é uma contradição, portanto, a suposição é falsa. ■

Corolário 5.2.2 $G^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$ *sse* $Goal_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Proposição 5.2.3 $\models Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Pela definição de I e B em 5.1.1, tem-se que $G^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$, então, existe algum $w'' \in B^w_t(\imath)$ tal que existe algum $w' \in I^w_t(\imath)$ onde $w' \sqsubseteq w''$. Agora assuma que para algum $\langle M, V, w_t\rangle$ qualquer, $\models Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Daí, conclui-se que $I^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$, $M, V, w_t \models Int_\imath(\varphi)$ e $M, V, w_t \models Bel_\imath(\neg Opt(\varphi))$. Pela semântica de Int, conclui-se que para todo $x' \in I^w_t(\imath)$, $M, V, x'_t \models Opt(\varphi)$. Todavia, pela semântica de Bel, sabe-se que para todo $y'' \in B^w_t(\imath)$, $M, V, w''_t \models \neg Opt(\varphi)$ e por consequência de $I^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$, existe pelo menos um $w' \in G^w_t(\imath)$ tal que $M, V, w'_t \models \neg Opt(\varphi)$. Portanto, há uma contradição. Logo, a suposição há de ser falsa. ■

Corolário 5.2.3 $I^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$ *sse* $Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Para finalizar, confirmar-se-á que a lógica BDI cujas relações de acessibilidade são definidas conforme 5.1.1 obedece aos dois princípios da tese de assimetria os quais são representados pela propriedade $Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Teorema 5.2.1 *Se* $Int_\imath(Opt(\varphi))$ *então* $\neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Prova trivial por silogismo hipotético, a partir das proposições 5.2.1 e 5.2.2. ■

Teorema 5.2.2 $I^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$

PROVA: Prova trivial a partir do teorema 5.2.1 e da proposição 5.2.3. ■

5.2.3 Caracterização de Crenças, Desejos e Intenções

As definição das relações de acessibilidade apresentadas na lógica BDI, criada neste capítulo, são mais gerais do que as de Wooldridge em [Woo00a], pois aqui é considerado além das relações provenientes de um dado agente (como são definidas as relações de acessibilidade em [Woo00a]) também se permitem relações provenientes de um conjunto de agentes. Portanto, seja $\overline{\mathcal{M}}$ um conectivo modal e $R'_{\overline{\mathcal{M}}}$ a relação de acessibilidade referente à $\overline{\mathcal{M}}$. Então se $\overline{\mathcal{M}}_i(\varphi)$ é satisfazível em w_t, logo haverá um mundo de w' relacionado a w_t via $R'_{\overline{\mathcal{M}}}$ no qual φ é satisfeito no tempo t. E $B_t^w(\imath)$, $G_t^w(\imath)$ e $I_t^w(\imath)$ são definidos conforme $B_t^w(i)$, $G_t^w(i)$ e $I_t^w(i)$, apresentados na subseção 4.1.4.

Bem como é explicitado por Rao e Georgeff em [RG91, RG93]: permite-se nesta lógica intenções, objetivos e crenças sob qualquer fórmula. Consequentemente, é possível haver intenções sobre crenças, sobre objetivos e até intenções sobre intenções, enfim, o agente pode ter qualquer tipo de intenção, no entanto, ele só age sobre a última delas.

5.3 Teoria Formal da Nova Lógica BDI

Apresentar-se-á nesta seção a teoria formal da lógica BDI cuja abordagem semântica foi apresentada na seção anterior, denotada por T^{3M} . Ver-se-á que todos os axiomas do sistema BDI básico são facilmente prováveis semanticamente, sendo o contrário também verdadeiro (ou seja, a abordagem semântica também é assegurada pela teoria formal); o que, apesar de não provada, há a intuição de que tal teoria formal da lógica BDI, a ser definida nesta seção, é completa e correta.

A linguagem desta teoria formal BDI será a linguagem $\mathcal{L}^{3M'} = \langle \Sigma^{3M'}, \mathcal{G}^{3M'} \rangle$, onde $\Sigma_{3M'} = X \cup \Sigma_d \cup \Sigma_P \cup Ling(\mathcal{L}^E) \cup Ling\mathcal{L}^{Gr}$, $\Sigma_{L3M'} = \Sigma_{L^E} \cup \{\neg, \rightarrow, \forall, \bigcirc, \Diamond, ?, \mathcal{U},$ $Inev, Bel, Goal, Int, Fail, \}$ e $\mathcal{G}^{3M'} = \mathcal{G}^{3M} -$ $\{F_3, F_4, F_6, F_7, F_9, F_{13}, F_{18}, F_{23}, F_{25}, F_{26}\}$. Esta linguagem pode ser vista como uma simplificação da linguagem $\mathcal{L}^{3M}$, pois, é construída com um número reduzido de operadores lógicos e, consequentemente, um número reduzido de regras gramaticais.

Observação 5.3.1 *Os teoremas envolvendo operadores não definidos pelas regras gramaticais antes explicitadas podem ser provados usando as respectivas equivalências:*

1. $\varphi \vee \phi \equiv \neg\varphi \rightarrow \phi$

2. $\varphi \wedge \phi \equiv \neg(\varphi \rightarrow \neg\phi)$

3. $\varphi \leftrightarrow \phi \equiv (\varphi \rightarrow \phi) \wedge (\phi \rightarrow \varphi)$

4. $\exists x.\varphi \equiv \neg\forall x.\neg\varphi$

5. $\Box\varphi \equiv \neg\Diamond\neg\varphi$

6. $\varphi \mathcal{W} \psi \equiv (\varphi \mathcal{U} \psi) \vee \Box \varphi$

7. $Success_\imath(\alpha, [\varphi]) \equiv \neg Fail_\imath(\alpha, [\varphi])$

8. $Execute_\imath(\alpha, [\varphi]) \equiv Success_\imath(\alpha, [\varphi]) \vee Fail_\imath(\alpha, [\varphi])$

9. $Opt\varphi \equiv \neg Inev \neg \varphi$

10. $Group(gr)(e) \equiv \Diamond Execute_{gr}(e)$

A relação B é uma relação serial, transitiva e euclidiana, garantindo, assim, que dada uma situação, sempre haverá um mundo de crença, que os mundos de crenças se inter-relacionam de forma transitiva e euclidiana. Assim, a axiomatização para as crenças é o sistema KD45 (S5), pois também são considerados os axiomas Δ da lógica clássica de 1ª ordem. Já os desejos e as intenções são relações de acessibilidade seriais (axioma D) e são fechados sob a implicação (axioma K), tendo, então, ambos, um sistema KD como axiomatização.

As regras de inferência desta teoria formal são as regras do MP, Gen, Nes, além da regra Nes aplicada aos operadores Bel, $Goal$, Int: NesB, NesG, e NesI; mostradas a seguir.

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash Bel_\imath(\varphi)} \tag{5.4}$$

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash Goal_\imath(\varphi)} \tag{5.5}$$

$$\frac{\Gamma \vdash \varphi}{\Gamma \vdash Int_\imath(\varphi)} \tag{5.6}$$

A inclusão das regras NesB, NesG e NesI significa que o agente crê, deseja e pretende todas as fórmulas válidas, o que acarreta no problema da onisciência lógica, primeiramente tratada em [Hin75], a qual foi resolvida por Vardi (com respeito à lógica epistêmica) em [Var86] através do tratamento de crenças e conhecimento como conjuntos de fórmulas. O problema da onisciência lógica é estendido para a lógica BDI da seguinte forma: se φ é uma fórmula válida (verdadeira em todos os estados de todos os caminhos de todos os mundos de todos os modelos), então, o agente crerá, desejará e pretenderá φ o que não é razoável para as semânticas de crença, objetivo e intenção. Pois, se o agente i crer em φ e $\varphi \rightarrow \phi$ é uma fórmula válida, então, ϕ também faria parte das crenças de i, já que também estaria no seu mundo de crenças; além disso, desejar uma fórmula válida e pretendê-la, é um problema de sobreposição de compromissos idênticos ("*overcommitment*") – como seria, por exemplo ter a intenção de que após o por do sol será noite? É possível também o seguinte exemplo: mesmo que É_VERMELHA($cadeira$) seja uma fórmula válida, pode ser o caso do agente i ser daltônico e Bel_i(É_VERDE($cadeira$)) ou, simplesmente, Bel_i ($\neg$É_VERMELHA($cadeira$)). Contudo, a generalização é necessária, pois se φ é uma fórmula válida, então, φ é verdadeira em todos os mundos e, por isso, ela também será verdadeira nos mundos de crenças, de objetivos e de intenções.

Os outros axiomas são referentes às inter-relações entre intenções, objetivos e crenças. Os axiomas escolhidos aqui não são os mesmos de [RG91, RG93], primeiramente, devido às discordâncias entre as inter-relações entre crenças e desejos e, consequentemente, entre crenças e intenções. Além de não se ver necessidade da definição de A7 e A6 como axiomas (o que será comentado com mais detalhes na seção a seguir).

Ver-se-á agora os axiomas e suas respectivas formalizações de condições semânticas.

[A1]$Goal_\imath\,(Opt(\varphi)) \rightarrow Bel_\imath\,(\neg Opt(\neg\varphi))$	[SC1]$\exists w', w''.(w' \in B_t^w(\imath) \wedge w'' \in G_t^w(\imath) \wedge w'' \sqsubseteq w')$
[A2]$Int_\imath\,(Opt(\varphi)) \rightarrow Goal_\imath\,(Opt(\varphi))$	[SC2]$\forall w'.(w' \in G_t^w(\imath) \rightarrow \exists w''.(w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq w')$
[A3]$Int_\imath(\varphi) \rightarrow Bel_\imath(Int_\imath(\varphi))$	[SC3]$\forall w', w''.(w' \in I_t^w(\imath) \rightarrow w'' \in B_t^w(\imath) \wedge w' \in I_t^{w''}(\imath))$
[A4]$Goal_\imath\,(\varphi) \rightarrow Bel_\imath(Goal_\imath(\varphi))$	[SC4]$\forall w', w''.(w' \in G_t^w(\imath) \rightarrow w'' \in B_t^w(\imath) \wedge w' \in G_t^{w''}(\imath))$
[A5]$Int_\imath\,(\varphi) \rightarrow Goal_\imath(Int_\imath(\varphi))$	[SC5]$\forall w', w''.(w' \in I_t^w(\imath) \rightarrow w'' \in G_t^w(\imath) \wedge w' \in I_t^{w''}(\imath))$

Os axiomas A1-A5 compõem o sistema BDI básico. Note que SC1 e SC2 são as inter-relações $G_t^w(\imath) \cap_{sub} B_t^w(\imath)$ e $G_t^w(\imath) \subseteq_{sup} I_t^w(\imath)$ cujas correspondências sintático-semântica entre elas e os axiomas A1 e A2 já foram provadas, respectivamente, pelas proposições 5.2.1 e 5.2.2. Já as correspondências entre os axiomas A3 e SC3, A4 e SC4 e A5 e SC5 são garantidas pela definição 5.2.1. Para isso, vide as proposições a seguir.

Proposição 5.3.1 $\models Int_\imath\,Opt(\varphi) \rightarrow Bel_\imath(Int_\imath\,Opt(\varphi))$.

PROVA: Assuma que para um modelo arbitrário $\langle M, V, w, t\rangle$, $M, V, w_t \models Int_\imath\,Opt(\varphi)$. A partir da semântica de Int, obtém-se que para todo $x' \in I_t^w(\imath)$ $M, V, x'_t \models Opt(\varphi)$. Pela definição 5.2.1, tem-se que $\forall x', y''.(x' \in I_t^w(\imath) \rightarrow y'' \in B_t^w(\imath) \wedge x' \in I_t^{y''}(\imath))$. Portanto, haverá pelo menos um mundo $w'' \in B_t^w(\imath)$ que se relaciona com os mundos de intenções pertencentes à $I_t^w(\imath)$. Por consequência disto, tem-se que $M, V, w''_t \models (Int_\imath\,Opt(\varphi))$ e, consequentemente, $M, V, w_t \models Bel\imath(Int_\imath\,Opt(\varphi))$ para o mesmo $\langle M, V, w, t\rangle$. Logo, $\models Int_\imath\,Opt(\varphi) \rightarrow Bel_\imath(Int_\imath\,Opt(\varphi))$. ■

Corolário 5.3.1 *Seja φ uma fórmula de $L^{3M'}$. $\forall w', w''.(w' \in I_t^w(\imath) \rightarrow w'' \in B_t^w(\imath) \wedge w' \in I_t^{w''}(\imath))$ sse $Int_\imath\,Opt(\varphi) \rightarrow Bel_\imath(Int_\imath\,Opt(\varphi))$.*

Proposição 5.3.2 $\models Goal_\imath\,Opt(\varphi) \rightarrow Bel_\imath(Int_\imath\,Opt(\varphi))$.

PROVA: Prova análoga à proposição 5.3.1. ■

Corolário 5.3.2 *Seja φ uma fórmula de $L^{3M'}$. $\forall w', w''.(w' \in G_t^w(\imath) \rightarrow w'' \in B_t^w(\imath) \wedge w' \in G_t^{w''}(\imath))$ sse $Goal_\imath\,Opt(\varphi) \rightarrow Bel_\imath(Int_\imath\,Opt(\varphi))$.*

Proposição 5.3.3 $\models Int_\imath\,Opt(\varphi) \rightarrow Goal_\imath(Int_\imath\,Opt(\varphi))$.

PROVA: Prova análoga à proposição 5.3.1. ■

Corolário 5.3.3 *Seja φ uma fórmula de $L^{3M'}$. $\forall w', w''.(w' \in I_t^w(\imath) \rightarrow w'' \in G_t^w(\imath) \wedge w' \in I_t^{w''}(\imath))$ sse, $Int_\imath\, Opt(\varphi) \rightarrow Goal_\imath(Int_\imath\, Opt(\varphi))$.*

Os axiomas seguintes são relacionados às estratégias de compromisso do agente. Portanto, o sistema BDI pode ser definido pelo sistema básico mais um (ou dois) dos axiomas apresentados em seguida.

[A6a]$(Int_\imath(Inev(\Diamond\varphi)) \rightarrow Int_\imath\,(Inev(\Diamond\varphi))\mathcal{U}\,Bel_\imath(\varphi?))$

[A6b]$Int_\imath(Inev(\Diamond\varphi)) \rightarrow Int_\imath(Inev(\Diamond\varphi))\mathcal{U}\,Bel_\imath(\varphi?) \vee (Bel_\imath(\neg Opt(\Diamond\varphi)))$

[A6c]$Int_\imath(Inev(\Diamond\varphi)) \rightarrow Int_\imath(Inev(\Diamond\varphi))\mathcal{U}\,(Bel_\imath(\varphi?)) \vee (\neg Goal_\imath(Opt(\Diamond\varphi)))$

Esses axiomas representam, respectivamente, as estratégia de compromisso cego, simples e mente-aberta e são bastante similares aos axiomas AI10a, AI10b and AI10c de [RG93] e AI9a, AI9b e AI9c de [RG91] explicitados como A10a, A10b e A10c na subseção 4.2.3.

Observação 5.3.2 *O compromisso poderia ser mantido até que φ? simplesmente ao invés de $Bel_\imath(\varphi?)$ como é definido. Entretanto, Bratman ao tratar as reconsiderações de intenções acha relevante o que o agente acredita sobre a situação a qual se encontra, logo, o agente deve acreditar que a intenção foi satisfeita para que o mesmo possa abandoná-la.*

Em A6a, o agente ou grupo de agentes mantém a intenção de realizar inevitavelmente algo até que isto tenha sido realmente satisfeito. Na estratégia de compromisso simples A6b, o(s) agente(s) mantém sua intenção até que ela seja satisfeita ou até que ele(s) acredite(m) que não é mais possível satisfazê-la. E na estratégia mente-aberta A6c ele(s) mantém a intenção enquanto há o desejo de satisfazê-la.

Note que as formalizações das estratégias de compromisso são semelhantes às definidas em [RG91, RG93], mas deixa clara a satisfação da condição ou da ação φ através do operador '?' e deixa claro também que o ponto de parada é a crença (nos casos de A6b) na impossibilidade de satisfação de φ.

5.4 Considerações Finais

Este capítulo inicia apresentando as motivações de se construir um novo sistema lógico BDI. Além das ambiguidades entre as propriedades sintáticas e as propriedades semânticas apresentadas em [RG91, RG93] e em [Woo00a] já referenciadas no capítulo anterior; há a motivação de formalizar um sistema mais condizente com a teoria intencional de Bratman [Bra87]. Foram, portanto, apresentados neste capítulo, sistemas lógicos BDI (abordagens formal e semântica) baseados nos sistemas propostos por Rao e Georgeff em [RG91, RG93] e por Wooldridge em [Woo00a] com algumas alterações de modo que a abordagem se tornasse mais clara[3], mais elegante, mais próxima à teoria intencional de Bratman [Bra87] e, consequentemente, mais aplicável à realidade de agentes inteligentes humanos. As mudanças introduzidas pelo sistema formal BDI definido neste capítulo ainda não soluciona problemas de onisciência lógica e de efeitos indesejados, mas ainda, se caracteriza como uma formalização da teoria de raciocínio prático de agentes humanos desenvolvida por Bratman em

[3]Mais clara, pois houve a preocupação de definir as e demonstrar relações sintáticas e semânticas entre as atitudes mentais cujos entendimentos eram bastante nebulosos tanto no livro de Wooldridge [Woo00a] quanto nos artigos de Rao e Georgeff [RG91, RG93].

[Bra87]. Porque mantém a postura intencional e define o raciocínio prático de agentes racionais com base nas suas crenças, desejos e intenções, onde tais atitudes possuem a mesma importância não podendo ser substituído uma pela outra.

O problema da onisciência lógica, segundo Cresswell em [Cre70], [Cre72] e [Cre73] seria supostamente resolvido através de permissão de mundos não-clássicos na semântica da lógica modal, no entanto, em [Var86] argumenta-se que mesmo a semântica sendo não-clássica, uma fórmula válida continuaria sendo verdadeira em todos os mundos, ou seja, o agente continuaria acreditando em, desejando e pretendendo: as "consequências lógicas" de suas(eus) crenças, desejos e intenções (respectivamente); e as fórmulas satisfeitas em seus mundos de crença, desejo e intenção respectivamente.

Outros aspectos da teoria de Bratman não foram abordados aqui, tais como aspectos do comportamento do agente referente à reconsideração de intenções ou à própria deliberação de planos ao haver mudança no ambiente ou nas circunstâncias. Entretanto, com tal sistema é possível categorizar diversos tipos de agentes cognitivos, suas diferentes estratégias de compromisso e captura do processo de revisão de crenças, desejos e intenções, bem como, fornece uma lógica base para ser fuzzyficada. Esta, por sua vez, proverá um sistema de inferência BDI fuzzy capaz de formalizar a deliberação e reconsideração de crenças.

Capítulo 6

Teoria Fuzzy

A teoria fuzzy pode ser subdividida em teoria dos conjuntos fuzzy, a lógica fuzzy e raciocínio fuzzy. Aqui estudar-se-á esses dois primeiros componentes para tornar mais clara a visualização da nova proposta de lógica BDI a qual será apresentada no sétimo capítulo.

6.1 Introdução

A lógica fuzzy foi criada por Zadeh em 1965 a partir do artigo [Zad65] e seu estudo foi continuado por Cox, Kasabov, Terano, Kosko, Tsoukalas, Watannabe, Hájek, entre vários outros; implodindo em duas vertentes: a lógica fuzzy em "*Broad Sense*" que tem por objetivo principal desenvolver sistemas baseados em raciocínios fuzzy (tais como os controladores fuzzy); e a lógica fuzzy em "*Narrow Sense*" que trata a mesma como uma lógica simbólica e, portanto, estuda aspectos tais como suas teorias formais, formas normais, etc. Recentemente, consideráveis progressos têm sido alcançados em relação aos aspectos matemáticos (formal, simbólico) da lógica fuzzy como uma lógica com noção comparativa da verdade.

6.2 Teoria dos Conjuntos Fuzzy

A teoria dos conjuntos fuzzy tem por distinção a forma de definir a pertinência de um elemento em um conjunto através de um valor dentro do intervalo real [0,1], ou seja, além da atribuição clássica dos valores 0 e 1, que indicam se um determinado elemento pertence (valor 1) ou não (valor 0) a um determinado conjunto, essa teoria determina em que medida um elemento pertence ou não a um determinado conjunto através de um grau chamado de grau de pertinência (ou grau de verdade ou grau de certeza). Então, seja o conjunto fuzzy A, o elemento x e $\mu_A(x)$ (o grau de pertinência de x no conjunto A); tem-se que se $\mu_A(x) = 1$, então, x **é um elemento do conjunto** A; se $\mu_A(x) = 0$, então x **não é um elemento do conjunto** A; e se $1 > \mu_A(x) > 0$, então, **o elemento pertence, mas não totalmente ao conjunto** A.

Por isso, as teorias dos conjuntos fuzzy e, consequentemente, as lógicas fuzzy, isto é, as lógicas as quais são baseadas numa teoria de conjuntos nebulosos, tornam-se uma ferramenta bastante útil para representar o conhecimento incerto (raciocínio aproximado) o qual é, de fato, mais adequado para representar situações da nossa realidade.

Um exemplo de como conjuntos clássicos e fuzzy podem ser usados é visto a seguir: o exemplo ilustra como um conjunto clássico e um conjunto fuzzy poderiam determinar se uma certa temperatura é fria.

A representação do problema utilizando um conjunto clássico trataria o problema da seguinte forma:

- T é o conjunto de todas as temperaturas.
- F é o conjunto das temperaturas que são frias e é definido segundo a equação 6.1. Podendo também ser representado pelo plano cartesiano da figura 6.1.

$$F = \{x \in T \mid 0 \leq x \leq 10\} \tag{6.1}$$

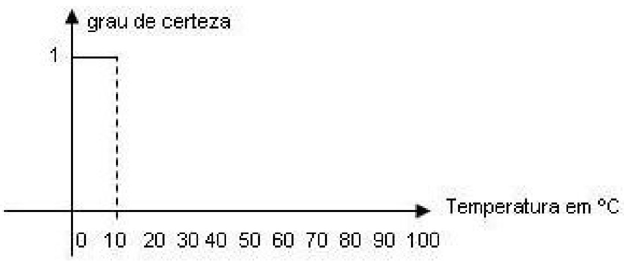


Figura 6.1: Conceito clássico de frio.

Já em conjuntos fuzzy, o problema é mapeado da seguinte forma: a função de pertinência é definida por $\mu_F(x) : U \rightarrow [0, 1]$, em que U é o universo de discurso do conjunto fuzzy F e $\mu_F(x)$ é o grau de pertinência definido no intervalo [0,1]. Podendo ser representado pelo gráfico na figura 6.2.

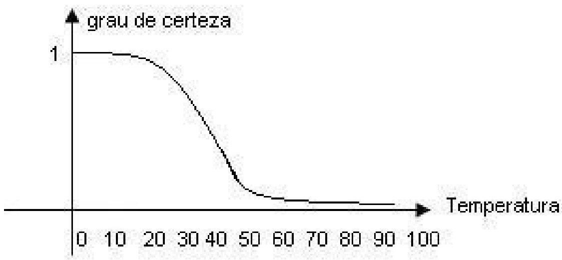


Figura 6.2: Conceito difuso de frio.

Observe que na representação clássica a temperatura $10,1°$ já é considerado não frio. Enquanto na representação fuzzy a mesma temperatura é fria com grau próximo a 1. Dessa forma, nota-se porque os conjuntos fuzzy são mais indicados para representar situações do cotidiano. No gráfico anterior observa-se que cada valor de temperatura está ligado a um valor do grau de certeza que determina quanto tal temperatura é considerada fria.

6.2.1 Representação dos Conjuntos Fuzzy e Funções de Pertinência

Os conjuntos fuzzy são geralmente representados como uma coleção de pares ordenados (equação 6.2). Mas também, podem ser escritos como um somatório (equação 6.3) – caso o universo de discurso seja discreto – ou como uma integral, caso o universo seja contínuo (equação 6.4).

$$A = \{(x, \mu_A(x)) \mid x \in U\} \tag{6.2}$$

onde U é o universo de discurso, x um elemento em U e A é o conjunto fuzzy definido em U;

$$A = \sum \mu_A(x_i)/x_i \tag{6.3}$$

$$\int_{U\mu_A(x)/x} \tag{6.4}$$

A função de pertinência é uma função que estabelece uma ligação entre um elemento do domínio e um valor verdade e indica o grau de pertinência do elemento no conjunto. O valor deste grau pertence ao intervalo real $[0, 1]$. Para se determinar o grau de pertinência de um determinado elemento é necessário que o especialista responsável por tal tarefa conheça bem o conjunto e seu significado. E analise qual será o melhor formato da função, pois a forma (os tipos de funções de pertinência) definirá o comportamento do conjunto fuzzy a ser representado, escolhendo assim o tipo de curva que melhor se aproxima do problema em questão.

6.3 Lógica Fuzzy

Após tratar da teoria dos conjuntos fuzzy, esta seção tem o objetivo de falar da lógica que se baseia em tal teoria. Ver-se-á, então, uma lógica que permite modelar sistemas bastante complexos, com capacidade de trabalhar incertezas, usando um alto nível de abstração, proveniente do conhecimento do engenheiro ou especialista, e uma linguagem simplificada baseada em variáveis linguísticas, modificadores linguísticos, proposições e regras de inferência.

6.3.1 Termos Linguísticos

As variáveis linguísticas são variáveis cujos valores são palavras ou sentenças, ou melhor, nomes de conjuntos fuzzy (que possuem um significado de acordo com o que representa). Deste modo, uma variável linguística pode assumir um dentre todos os valores de conjuntos fuzzy que a compõem com um determinado grau, tais conjuntos são chamados de termos da variável. Na figura 6.3 exemplifica-se uma variável linguística temperatura formada pelos termos "Frio", "Normal" e "Quente".

Variáveis linguísticas podem conter também modificadores (também linguísticos). Estes proporcionam uma mudança na função de pertinência de forma que ela relacione o valor semântico do modificador à uma equação matemática, fazendo com que o valor original da função de pertinência seja alterado.

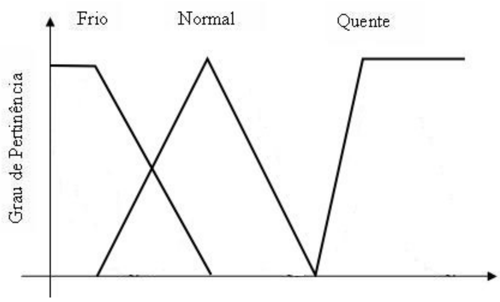


Figura 6.3: Variável linguística fuzzy temperatura.

Daí, seja A um conjunto fuzzy e $\mu_A(x)$ a função de pertinência que determina o grau de certeza de x em A. O modificador linguístico 'm' é definido de acordo com a equação 6.5.

$$\mu_{mA}(x) = f(\mu_A(x)) \tag{6.5}$$

Vê-se dois exemplos de modificadores linguísticos nas equações 6.6 e 6.7, representando, respectivamente, *não* e *pouco* aplicados ao conjunto "Frio" (já definido anteriormente):

não: Sendo $f_{\text{não}}(x) = 1 - x$ então

$$\mu_{\text{não}F}(x) = 1 - \mu_F(x) \tag{6.6}$$

muito: Sendo $f_{pouco}(x) = x^2$ então

$$\mu_{\text{pouco}F}(x) = (\mu_F(x))^2 \tag{6.7}$$

6.3.2 Relações Fuzzy

Com a criação da teoria dos conjuntos, surge a necessidade de relacionar os elementos de tais tipos de conjuntos para variados fins. As relações são associações entre elementos de dois ou mais conjuntos e retorna um grau de verdade cujo valor está no intervalo real [0,1]. Em alguns casos, esses graus só podem assumir valores clássicos, como pode ser observados em relações do tipo "*é dividido por*", mas há também infinitas formas de estender o grau de verdade para valores reais de [0,1] tal como determinar o quão cheio estão os tanques, quão quente está certa mistura de substâncias e, principalmente, quando as relações se submetem a propor comparações, onde o fator *Interpretação do Especialista* é inerente ao problema.

As relações fuzzy podem ser definidas em Produtos Cartesianos, linguisticamente, através de grafos, entre outras formas. Mas, geralmente, são definidas como um conjunto de pares segundo a equação 6.8 ou listando todos os elementos da relação conforme a equação 6.9 (no caso de uma relação binária R definida em $X \times Y$):

$$R = \{((x,y), \mu_R(x,y)) | x \in X \, e \, y \in Y\} \tag{6.8}$$

$$\begin{aligned} R = \{&((x_1,y_1), \mu_R(x_1,y_1)), ((x_1,y_2), \mu_R(x_1,y_2)), ((x_1,y_3), \mu_R(x_1,y_3)), \\ &((x_2,y_1), \mu_R(x_2,y_1)), \ldots\} \end{aligned} \tag{6.9}$$

6.3.3 Semântica Fuzzy

Na lógica clássica, uma proposição ou é verdadeira ou é falsa, já na lógica fuzzy, uma proposição tem um certo grau de verdade ou um certo grau de falsidade. Para fazer a composição de duas proposições usa-se os operadores ou conectivos lógicos. E estes podem ser aplicados em proposições fuzzy e em conjuntos fuzzy sendo o grau de verdade da proposição obtido através da função de pertinência.

As normas triangulares (ou simplesmente t-normas) e seus duais as conormas triangulares (t-conormas) que antes eram usadas como medidores de distâncias em espaços métricos probabilísticos, atualmente, são aplicadas como conectivos lógicos fuzzy ([ATV80]). Como pode ser usada qualquer t-norma, é mais interessante deixar o especialista decidir qual a melhor t-norma que se aplica a seu problema. Portanto, não será adotado nenhum operador fuzzy específico para este trabalho, serão definidas apenas as propriedades as quais eles deverão ter.

A sintaxe desta lógica fuzzy é constituída pela mesma linguagem $\mathcal{L}^P$ da lógica clássica de 1ª ordem, no entanto, a semântica clássica agora é substituída pela semântica fuzzy, pois será dada uma nova interpretação aos elementos de Σ_L e às fbf de L^P. Para isso, ver-se-á agora uma classe de extensões fuzzy de conectivos, as quais podem ser encontradas em [Bac04, BBS03, Fod91, HN02, KN99, Les03, Nav99, SS63], sendo o da bi-implicação encontrado apenas em [BC08] e [Roc07].

Seja $\mathbb{I} = [0, 1]$. Uma operação binária T em $\mathbb{I}$, isto é, $T : \mathbb{I} \times \mathbb{I} \longrightarrow \mathbb{I}$ é uma t-norma, se tiver as propriedades:

Simetria: $\mathrm{T}(x, y) = \mathrm{T}(x, y), \forall x, y \in \mathbb{I}$,

Associatividade: $\mathrm{T}(x, \mathrm{T}(y, z)) = \mathrm{T}(\mathrm{T}(x, y), z), \forall x, y, z \in \mathbb{I}$,

Monotonicidade: Se $x \leq x'$ e $y \leq y'$ então $\mathrm{T}(x, y) \leq \mathrm{T}(x', y'), \forall x, x', y, y' \in \mathbb{I}$ e

1-Identidade: $\mathrm{T}(x, 1) = x, \forall x \in \mathbb{I}$.

Como é conhecido, qualquer t-norma estende a conjunção clássica, no sentido de que T restringido aos valores booleanos se torna a operação de multiplicação booleana.

Há uma quantidade infinita de t-normas. E as mais usuais são as t-normas de Gödel (G) e do Produto (P), definidas por $G(x, y) = \min\{x, y\}$ e $P(x, y) = xy$, respectivamente.

Seja S uma operação binária sobre $\mathbb{I}$. S é uma t-conorma, se a mesma for simétrica, associativa, monotônica e $\mathrm{S}(x, 0) = x, \forall x \in \mathbb{I}$ (0-Identidade). Um exemplo típico de t-conorma é $\mathrm{S}(x, y) = \max(x, y)$.

Para caracterizar a Negação, tem-se: seja N uma operação unária sobre $\mathbb{I}$. N é uma "negação fuzzy" se a operação unário é:

- preserva limites, ou seja, $\mathrm{N}(0) = 1$ e $\mathrm{N}(1) = 0$
- antitônica, isto é, se $x \geqslant y$ então $\mathrm{N}(x) \leqslant \mathrm{N}(y)$.

Uma negação fuzzy N é dita forte se satisfazer também a propriedade involutiva: $\mathrm{N}(\mathrm{N}(x)) = x$, $\forall x \in \mathbb{I}$ – segundo [BBS03] e [KMP00]. Um exemplo típico da negação fuzzy forte é $\mathrm{N}(x) = 1 - x$.

Na literatura, há muitas definições de implicações fuzzy, como por exemplo [Bac04, BBS03, Fod91, FR94, HN02, Les03, RK93, Yag83, Yag04]. Em tais definições, há o consenso de que as implicações fuzzy devem ter o mesmo comportamento que a implicação clássica para os casos crisp (valores extremos '0' e '1'). Aqui serão consideradas as propriedades estabelecidas por Fodor e Roubens em [FR94]. Seja J uma operação binária sobre $\mathbb{I}$. J é uma implicação fuzzy se ela satisfizer as seguintes propriedades $\forall x, y, z \in \mathbb{I}$,

I1: Se $x \leq z$ então $\mathrm{J}(x, y) \geq \mathrm{J}(z, y)$,

I2: Se $y \leq z$ então $\mathrm{J}(x, y) \leq \mathrm{J}(x, z)$,

I3: $\mathrm{J}(0, y) = 1$,

I4: $\mathrm{J}(x, 1) = 1$,

I5: $\mathrm{J}(1, 0) = 0$.

Seja B uma operação binária sobre $\mathbb{I}$. Segundo [Roc07, BC08], B é uma bi-implicação fuzzy se satisfizer as seguintes propriedades $\forall x, y, z \in \mathbb{I}$,

B1: $\mathrm{B}(x, y) = \mathrm{B}(y, x)$,

B2: Se $x = y$ então $\mathrm{B}(x, y) = 1$,

B3: $\mathrm{B}(0, 1) = 0$,

B4: Se $x \leq y \leq z$ então $\mathrm{B}(x, y) \geq \mathrm{B}(x, z)$,

B5: Se $x \leq y \leq z$ então $\mathrm{B}(y, z) \leq \mathrm{B}(x, z)$.

Assim, dada uma interpretação I e uma atribuição ρ a uma fórmula φ, diz-se que φ possui um modelo numa interpretação I ($I \models_\rho \varphi$) se para alguma atribuição ρ das variáveis livres, $V_\rho(\varphi) = 1$. Entretanto, nesta semântica V é redefinida como uma função que mapeia fbf de L^P em $\mathbb{I}$, denotado por $V : L^P \to \mathbb{I}$.

As interpretações das fórmulas moleculares obtidas pelas regras $F_2, \ldots, F_6$ já foram dadas anteriormente. E a valoração dos termos mantém a definição exposta na sub-seção 3.2.2. Já a definição das interpretações dos predicados e dos quantificadores é semelhante à feita em [Háj98]:

Definição 6.3.1 *Seja $\varphi \in \mathcal{L}^P$, $\mathcal{F} = \langle \mathrm{T}, \mathrm{S}, \mathrm{N}, \mathrm{J}, \mathrm{B} \rangle$ uma semântica fuzzy para os conectivos lógicos e ρ uma atribuição das variáveis livres, então:*

$V_\rho^{\mathcal{F}}(R(\tau_1, \ldots, \tau_n)) = \mu_R(V_\rho(\tau_1), \ldots, V_\rho(\tau_n))$.

$V_\rho^{\mathcal{F}}(\neg\varphi) = \mathrm{N}(V_{rho}\varphi)$

$V_\rho^{\mathcal{F}}(\varphi \vee \phi) = \mathrm{S}(V_\rho(\varphi), V_\rho(\phi))$

$V_\rho^{\mathcal{F}}(\varphi \wedge \phi) = \mathrm{T}(V_\rho(\varphi), V_\rho(\phi))$

$V_\rho^{\mathcal{F}}(\varphi \rightarrow \phi) = \mathrm{J}(V_\rho(\varphi), V_\rho(\phi))$

$V_\rho^{\mathcal{F}}(\varphi \leftrightarrow \phi) = \mathrm{B}(V_\rho(\varphi), V_\rho(\phi))$

$V_\rho^{\mathcal{F}}(\forall x.(\varphi)) = inf\{V_{\rho'}^{\mathcal{F}}(\varphi) \ | \ \rho'(y) = \rho(y)\, \forall y.(y \neq x)\}$, *onde* $\rho' : X \rightarrow D$.

$V_\rho^{\mathcal{F}}(\exists x.(\varphi)) = sup\{V_{\rho'}^{\mathcal{F}}(\varphi) \ | \ \rho'(y) = \rho(y)\, \forall y.(y \neq x)\}$, *onde* $\rho' : X \rightarrow D$.

Quando a semântica fuzzy estiver clara do contexto, omitiremos ela da função de valoração, ou seja em vez de $V_\rho^{\mathcal{F}}$ escreveremos simplesmente V_ρ.

A partir de qualquer t-norma é sempre possível obter canonicamente qualquer outro conectivo fuzzy.

Definição 6.3.2 *Seja* T *uma t-norma. Define-se:*

- $\mathrm{S_T}T(x, y) = \mathrm{N_T}(\mathrm{T}(\mathrm{N_T}(x), \mathrm{N_T}(y)))$
- $\mathrm{J_T}(x, y) = \sup\{z \,/\, \mathrm{T}(x, z) \leq y\}$
- $\mathrm{B_T}(x, y) = \mathrm{T}(\mathrm{J_T}(x, y), \mathrm{J_T}(y, x))$
- $\mathrm{N_T}(x) = I_{\mathrm{T}}(x, 0)$

A partir da linguagem $\mathcal{L}^P$ e de uma semântica fuzzy $\mathcal{F}$ dos conectivos proposicionais, define-se a consequência semântica fuzzy e a lógica fuzzy.

Definição 6.3.3 *Seja* $\varphi \in L^P$ *e então* φ *é verdadeira dada uma interpretação* I *e uma atribuição* ρ *para as variáveis livres, denotado por* $I \models_\rho^K \varphi$*, sse* $V_\rho^{\mathcal{F}}(\varphi) \geq k$ *tal que* k *é uma constante veritativa dada pelo sistema,* $k \in (0, 1]$ *e os graus de verdade maiores ou iguais a ela são considerados verdadeiros. Quando* $V_\rho^{\mathcal{F}}(\varphi) \geq k$*,* (I, ρ) *é dito um modelo de* φ*.*

Baseado nesta definição, uma fórmula $\varphi \in L^P$, dada uma interpretação I e uma atribuição ρ às variáveis, é dita satisfazível se existe algum modelo; caso contrário, diz-se que ela é insatisfazível. E uma fórmula φ é falsa para uma interpretação I e uma atribuição ρ, denotado por $I \not\models_\rho^K \varphi$, se $V_\rho(\varphi) < k$. Uma fórmula é dita verdadeira em uma interpretação I, se para qualquer ρ, $I \models_\rho^K \varphi$, isto é: $\forall \rho.(V_\rho(\varphi) \geq k)$ e é uma contradição, denotado por $I \not\models^K \varphi$, se para qualquer ρ, $V_\rho(\varphi) < k$. E φ é contingente em I, se existem atribuições ρ e ρ' tal que $I \models_\rho^K \varphi$ e $I \not\models_{\rho'}^K \varphi$.

Por fim, $\varphi \in L^P$ é universalmente válida, uma K-tautologia, sse para todas as interpretações e atribuições ρ, $V_\rho(\varphi) \geq k$ cuja denotação é $\models^K \varphi$.

Definição 6.3.4 *Seja a linguagem de 1ª ordem $\mathcal{L}^P = \langle \Sigma, \mathcal{G} \rangle$ e $\Gamma \subseteq L^P$. φ é uma consequência semântica fuzzy de Γ, denotado por $\Gamma \models^K \varphi$ se todo modelo de Γ é um modelo de φ, o que quer dizer que se para toda fbf $\phi \in \Gamma$ $V_\rho(\phi) \geq k$, então $V_\rho(\varphi) \geq k$.*

Já a lógica fuzzy sobre esta linguagem $\mathcal{L}^P$ é a lógica Log^F, onde

$$Log^F = \{\varphi \in L^P \ \mid \models^K \varphi\}$$

que também pode ser definida por

$$Log^F = \langle L^P, \models^K \rangle$$

A teoria formal fuzzy não será apresentada, pois, a lógica fuzzy descrita aqui é suficiente para a construção da lógica e da teoria formal BDI fuzzy.

6.4 Considerações Finais

Neste capítulo, poder-se-ia tratar sistemas de inferência fuzzy e detalhar mais os conceitos aplicados a este tema, como por exemplo, conjuntos e relações fuzzy. Entretanto, apesar de comentar a possibilidade de desenvolvimento, um sistema de inferência fuzzy não será construído no presente trabalho. Por isso, preocupa-se mais, por enquanto, em detalhar a descrição da semântica fuzzy a qual será utilizada na lógica BDI fuzzy.

Capítulo 7

Proposta de uma Lógica BDI Fuzzy

A proposta de fuzzificar a lógica BDI, tal como foi tratado no primeiro capítulo, tem algumas vantagens, principalmente, com respeito à habilidade de um agente ter dúvidas e distintos graus de vontades com relação às suas intenções e desejos. Em adição às já referenciadas vantagens, em alguns casos pontuais da teoria de Bratman [Bra87] há características nebulosas, como por exemplo em [Bra87], onde se descreve "desejos predominantes" em que i deseja predominantemente φ se i deseja φ estritamente mais do que i deseja realizar qualquer outra opção a qual i acha incompatível com relação a satisfazer φ; e em [Bra87], onde há a descrição de "*flat-out belief*" definidas como um certo tipo de aceitação. E apesar de no geral, Bratman tratar crenças, desejos e intenções conflitantes de forma clássica; isso leva a crer que Bratman, em [Bra87], já tinha pelo menos uma pequena idéia de que as atitudes mentais devem ser tratadas com um certo grau de "aceitação" o que formalmente seria a consideração de graus de verdade, não unicamente estrita. Contudo, a justificativa de que os seres humanos (cujas mentes nós conhecemos relativamente) possuem raciocínios, em sua maioria paraconsistentes – atitudes mentais conflitantes devem ser consideradas no raciocínio prático ou teórico sem que haja a trivialização do sistema e não podem ser simplesmente descartados em qualquer caso – e vivem num mundo, onde o raciocínio aproximado é bastante empregado. Isso é suficiente para que uma lógica que modela o raciocínio de agentes cognitivos (tal como a lógica BDI) seja fuzzyficada.

7.1 Introdução

A lógica BDI fuzzy, proposta aqui, é uma lógica modal fuzzy em que para cada fórmula, seja ela de estado ou de caminho, é atribuído um valor do intervalo [0,1] para determinar qual o seu grau de verdade.

Segundo Zhang, Sui, Cao e Guohua em [ZSCG06], o estudo das lógicas modais fuzzy foi iniciado por Hájek em [HH96] e [Háj98] onde foi fornecida uma axiomatização completa de um sistema S5 fuzzy. Godo e Rodríguez, em [GR02] e [RGG96] estendem a lógica de Hájek definindo as relações de acessibilidade como relações de similaridade fuzzy e em [ZSC04] Zhang, Sui e Cao desenvolve um sistema formal baseado na lógica modal proposicional fuzzy cujas propriedades são discutidas em [ZSCG06]. Esse artigo apresenta uma extensão fuzzy da lógica modal proposicional a partir da adição de uma asserção fuzzy $\langle\varphi, n\rangle$ à lógica modal proposicional, onde $n \in [0, 1]$ e cujo alfabeto é composto por operadores clássicos e modais de necessidade e possibilidade, um conjunto de símbolos proposicionais $VP = \{p_1, p_2, \ldots\}$ e cuja semântica de Kripke é construída através do modelo M=$\langle W, R, V\rangle$, onde W é um conjunto de mundos possíveis, R um conjunto de relações de acessibilidade sobre mundos e $V : W \times VP \rightarrow \{0, 1\}$. Assim $M, V, w \models \varphi$ se $V(w, \varphi) = 1$.

Com a adição da asserção fuzzy $\langle\varphi, n\rangle$, onde o número n é chamado de grau de credibilidade ("*believable degree*") de φ, sendo φ um símbolo proposicional. Daí é definida a satisfazibilidade de φ num determinado mundo possível pela seguinte regra:

$$Sat(w, \langle\varphi, n\rangle)\, sse\, V(w, \varphi) \geq n \tag{7.1}$$

Ou seja, nesta lógica modal proposicional fuzzy, diz-se que $\langle\varphi, n\rangle$ é satisfazível em w, denotado por $M, V, w \vDash \langle\varphi, n\rangle$ se existir um $w \in W$ tal que $Sat(w, \langle\varphi, n\rangle)$.

Esta foi a proposta de lógica modal proposicional fuzzy mais simples encontrada, a qual é bastante eficiente e é um possível caminho a ser tomado. No entanto, tal lógica não possui uma semântica muito bem definida, pois o valor de verdade n da proposição φ é dada pela asserção fuzzy (na sintaxe da lógica).

Em [Fit91], Melvin Fitting investiga duas famílias de lógica modal multi-valorada: uma que considera a valoração de uma fórmula necessária ou possível em relação ao mundo em que a sub-fórmula estará; e uma que valora as fórmulas (necessárias e possíveis), considerando o mundo e as relação de acessibilidade. Tais lógicas podem ser facilmente estendidas para uma lógica modal fuzzy.

A valoração de uma fórmula $\Box\varphi$ em um mundo w, denotado por $\upsilon(w, \Box\varphi)$ é definida pela equação 7.2 (segundo a versão I da lógica de Fitting) e pela equação 7.4 (segundo a versão II); enquanto $\upsilon(w, \Diamond\varphi)$ é definido por 7.3 e 7.5 (segundo versões I e II respectivamente).

$$\upsilon(w, \Box\varphi) = \bigwedge\{\upsilon(w', \varphi)/wRw'\} \tag{7.2}$$

$$\upsilon(w, \Diamond\varphi) = \bigvee\{\upsilon(w', \varphi)/wRw'\} \tag{7.3}$$

$$\upsilon(w, \Box\varphi) = \bigwedge\{R(w, w') \rightarrow \upsilon(w', \varphi)/wRw'\} \tag{7.4}$$

$$\upsilon(w, \Diamond\varphi) = \bigvee\{R(w, w') \rightarrow \upsilon(w', \varphi)/wRw'\} \tag{7.5}$$

As valorações são assim definidas, pois: caso $\phi = \Diamond\varphi$, então φ é satisfeita em pelo menos um dos mundos relacionados ao mundo atual e, por isso, o valor de ϕ é dado pela disjunção dos valores de φ nos mundos relacionados (o "maior" de todos esses valores); enquanto caso $\phi = \Box\varphi$, então φ é verdadeiro em todos os mundo relacionados, portanto ϕ deve ser definido como a conjunção dos valores de φ nestes mundos, assim, se φ for falso em algum destes, então, ϕ também será.

Estendendo as equações 7.2 e 7.3 para uma semântica modal fuzzy tem-se a forma de valoração dada em [ZSC04] e em [ZSCG06], onde $\upsilon_\rho : W \times L^P \to [0,1]$:

$$\upsilon(w, \Box\varphi) = inf\{\upsilon(w', \varphi)/wRw'\} \tag{7.6}$$

$$\upsilon(w, \Diamond\varphi) = sup\{\upsilon(w', \varphi)/wRw'\} \tag{7.7}$$

Observação 7.1.1 *Não nos preocuparemos em descrever agora a interpretação dos outros conectivos que por ventura existiria nesta linguagem. No entanto, pode-se supor que a valoração das outras fórmulas seriam definidas similarmente à semântica fuzzy descrita no capítulo sexto.*

A nova lógica BDI fuzzy a ser proposta aqui é uma extensão da lógica BDI desenvolvida no capítulo anterior. Havendo mudanças na semântica, pois se utilizará uma semântica fuzzy (a qual parte dela foi apresentada no quarto capítulo), e nas atitudes mentais consideradas, isto é, aqui será empregado desejo ao invés de objetivo, uma vez que não mais importa se os desejos são conflitantes, pois estar-se-á trabalhando com a semântica fuzzy que também é paraconsistente (vide [CB07] para mais detalhes).

7.2 Sintaxe da Lógica BDI Fuzzy

A linguagem da lógica BDI fuzzy, denotada por $\mathcal{L}^{MF}$, é constituída pelo alfabeto Σ^{MF}, onde

- $\Sigma_{FM} = X \cup \Sigma_d \cup \Sigma_{L_{FM}} \cup Ling(\mathcal{L}^E) \cup Ling(\mathcal{L}^{Gr}) \cup \Sigma_P$,
- $\Sigma_{L_{FM}} = \Sigma_{L'} \cup \{\bigcirc, \mathcal{U}, ?, Inev, Opt, Bel, Des, Int, Success, Fail, Execute, Group\}$ e
- $\mathcal{G}^{MF} = \mathcal{G}^{3M} - \{F_{13}, F_{15}\} \cup F_{15a}$.

Onde

$F_{15a} \quad \dfrac{\varphi,\ \imath}{Des_\imath\,(\varphi)}$

Devido à definição da linguagem ser similar a $\mathcal{L}^{3M}$, as definições acerca das fórmulas de estado e de caminho são mantidas e $Ling(\mathcal{L}^{MF}) = L^{MF}$.

A sintaxe é a mesma da nova lógica BDI desenvolvida no capítulo anterior deste trabalho, com exceção da substituição do conectivo $Goal$ pelo Des. Enquanto isso, há muitas possibilidades de construir uma semântica fuzzy para esta lógica. Primeiro, porque tanto os mundos quanto as atitudes mentais e a execução dos eventos podem ser fuzzyficados – para definir um mundo fuzzy basta aplicar uma semântica fuzzy aos conectivos temporais e clássicos; para estabelecer um raciocínio (cognição) fuzzy, as interpretações dos conectivos Bel, Des, e Int deverão ser interpretados segundo uma semântica fuzzy; e as ações são formalizadas de tal forma que se possa representar a não completa execução, sucesso ou falha de um evento se for definida uma semântica fuzzy aos operadores de ações. Segundo, porque tais semânticas podem ser definidas de duas formas através da aplicação de uma extensão fuzzy das semânticas modais fuzzy definidas por Melvin Fitting em [Fit91]. Para simplificar as apresentações de tais lógicas BDI fuzzy, mostrar-se-á apenas duas abordagens semânticas para a lógica BDI fuzzy: Uma (versão I) a qual utiliza a valoração das sentenças em função da satisfazibilidade das mesmas nos mundos; e outra (versão II) cuja valoração das sentenças são dadas através da satisfazibilidade das mesmas nos mundos e dos valores das relações – entre mundos (R') ou entre estados de tempo (R) o que dependerá do operador envolvido na sentença – pois essas são as lógicas mais genéricas, uma vez que consideram a fuzzyficação das entidades de ações, temporais, modais e clássicas.

Observação 7.2.1 *Que fique claro que a extensão da lógica multivalorada de Fitting não é a única forma de estabelecer uma lógica modal fuzzy. Comentários sobre o estado da arte das lógica modais fuzzy foram feitos na introdução deste capítulo.*

7.3 Semântica da Lógica BDI Fuzzy

As interpretações das fbf de L^{MF} são dadas pela semântica de Kripke onde os mundos são estruturas temporais discretas, ramificadas no futuro, de passado linear e cujas relações R são ainda totais e transitivas tal como foi definido no capítulo anterior para Log^{3M}. Entretanto, haverá diferenças nas definições das relações R e R', bem como das funções Suc_w e Fal_w com respeito à lógica BDI fuzzy versão II. Por isso, a semântica de cada lógica será definida em particular.

7.3.1 Semântica da Lógica BDI Fuzzy Versão I

Definição 7.3.1 *Um modelo M é uma tupla $\langle W, T, \boldsymbol{D}, R, B, D, I, \Phi, E\rangle$ onde W é o conjunto de mundos possíveis, T é o conjunto dos estados de tempo, R é um arco que relaciona dois estados de tempos, **D** é o universo de discurso que contém os domínios de ações (D_{Ac}), agentes (D_{Ag}) e grupos de agentes (D_{Gr}), grupos (D_{Gr}) e outros indivíduos (D_U), B, D e I são relações – de crença, desejo e intenção, respectivamente, entre mundos ($B \subseteq W \times W$, $D \subseteq W \times W$ e $I \subseteq W \times W$). Φ é o mapeamento das entidades de 1ª ordem aos elementos de **D**. Por fim, E é um conjunto enumerável de eventos.*

As relações de acessibilidade B D e I são definidas similarmente a B, G e I da equação 5.1.1. Daí, tem-se a seguinte definição de relações de acessibilidade:

Definição 7.3.2 *Sejam* $\begin{array}{l} B : L_{Gr} \rightarrow \wp\{W \times T \times W\} \\ D : L_{Gr} \rightarrow \wp\{W \times T \times W\} \\ I : L_{Gr} \rightarrow \wp\{W \times T \times W\} \end{array}$ *onde B é uma relação de acessibilidade de crença, D é uma relação de acessibilidade de desejo e I é uma relação de acessibilidade de intenção. $\langle B, D, I\rangle$ são relações de acessibilidade* BDI *do tipo I se satisfazem as seguintes condições:*

- $\exists w', w''.(w' \in B_t^w(\imath) \wedge w'' \in D_t^w(\imath) \wedge w'' \sqsubseteq w')$;

- $\forall w'.(w' \in D_t^w(\imath) \rightarrow \exists w''.(w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq w')$; *e*

- $\exists w', w''.(w' \in B_t^w(\imath) \wedge w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq w')$.

Quanto aos mundos e submundos, eles são definidos conforme as descrições 2 e 3 da subseção 5.2.2.

Como nesta lógica não há graduação de R nem R', então, a interpretação dos conectivos de ação permanecerão os mesmos (com graus estritos de valoração), enquanto a valoração das fórmulas serão estendidas a um grau fuzzy, como segue.

Definição 7.3.3 *V é uma função de valoração que mapeia uma fbf de uma situação (w, t) a um valor de [0,1], ou mais formalmente: $V : S_w \times L^{MF} \rightarrow [0, 1]$. Portanto, seja $V_t^w(\varphi)$ uma valoração da fbf $\varphi \in L^{MF}$.*

A atribuição dos valores verdade das variáveis (livres) são atribuídos em cada situação (w, t) por $\rho : W \times T \times Ling(\mathcal{L}^T) \rightarrow D_{w_t}$[1], sendo D_{w_t} o domínio (universo de discurso) da situação (w, t) tal que:
$\forall f \in \Sigma_F,\ w_t(f) : D_{w_t}^{aridade(f)} \rightarrow D_{w_t}$
$\forall P \in \Sigma_R,\ w_t(P) \subseteq D_{w_t}^{aridade(P)}$
$\forall a \in \Sigma_C,\ w_t(a) \in D_{w_t} \rho_{w_t}(a) = w_t(a)$ *para cada* $a \in \Sigma_C$
$\rho_{w_t}(x) = w_t(x)$ *para cada* $x \in X$
$\rho_{w_t}(f(\tau_1, \ldots \tau_n)) = w_t(f)(\rho_w(\tau_1), \ldots, \rho_w(\tau_n))$.

Os símbolos de relações são interpretados como relações fuzzy, os conectivos proposicionais tem uma semântica fuzzy, bem como, os símbolos modais e temporais, com exceção dos símbolos $Success$, $Fail$ e $Execute$. As relações de acessibilidade entre mundos e entre instantes de tempo não tem valoração fuzzy. Portanto, a definição dos conectivos da linguagem BDI *fuzzy versão I segue a interpretação abaixo:*[2]

$V_t^w(R(t1, \ldots, tn)) = \mu_R(V_t^w(\tau_1), \ldots, V_t^w(\tau_n))$.

$V_t^w(\neg\varphi) = \mathrm{N}(V_t^w(\varphi))$

$V_t^w((\varphi \wedge \phi)) = \mathrm{T}(V_t^w(\varphi), V_t^w(\phi))$,

$V_t^w((\varphi \vee \phi)) = \mathrm{S}(V_t^w(\varphi), V_w^t(\beta))$

$V_t^w((\varphi \rightarrow \phi)) = \mathrm{J}(V_t^w(\varphi), V_t^w(\phi))$,

$V_t^w((\varphi \leftrightarrow \phi)) = \mathrm{B}(V_t^w(\varphi), V_t^w(\phi))$,

$V_t^w(\forall x.(\phi)) = inf\{V_t^w(\phi)\ /\ V'^{\,w}_{\,t}(y) = V_t^w(y)\, \forall y.(y \neq x)\}$

$V_t^w(\exists x.(\phi)) = sup\{V_t^w(\phi)\ /\ V'^{\,w}_{\,t}(y) = V_t^w(y)\, \forall y.(y \neq x)\}$

$V_t^w(\Diamond\varphi) = sup\{V_{t'}^w(\varphi)\, /\, tRt'\}$,

$V_t^w(\Box\varphi) = inf\{V_t^{w'}(\varphi)\, /\, tRt'\}$,

[1]A linguagem dos termos $Ling(\mathcal{L}^T)$ é definida similarmente como foi definida na seção 3.2.2.

[2]As considerações feitas quanto a ρ na seção 3.4.1 são seguidas de forma análoga nesta semântica.

$V_t^w(\bigcirc\varphi) = inf\{V_{t'}^w(\varphi)\,/\,tRt' \text{ e } t' = c(1)\}$

$V_t^w(\varphi\mathcal{U}\,\phi) = min\{inf\{V_j^w(\varphi)\,/\,j > t \text{ and } j,t \in c\},\ V_{t'}^w(\phi)\}$, *onde*

$\quad t' = max\{j \in c\,/\ V_j^w(\neg\phi) \geq k\} + 1$

$V_t^w(Bel_\imath\varphi) = inf\{V_t^{w'}(\varphi)\ /\ wBw'\}$,

$V_t^w(Des_\imath\varphi) = inf\{V_t^{w'}(\varphi)\ /\ wDw'\}$,

$V_t^w(Int_\imath\varphi) = inf\{V_t^{w'}(\varphi)\ /\ wIw'\}$,

$V_t^w(Inev(\varphi)) = inf\{V_{t'}^w(\varphi)\ /\ t' \in CAMINHOS(w_t)\}$

$V_t^w(Opt(\varphi)) = sup\{V_{t'}^w(\varphi)\ /\ t' \in CAMINHOS(w_t)\}$

$$V_{tn}^w(Sucess_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in n \textit{ tal que } i < n \textit{ e } Suc_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_{tn}^w(Fail_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in n \textit{ tal que } i < n \textit{ e } Fal_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_{tn}^w(Execute_\imath(e)) = \begin{cases} 1 & \textit{, se existe } ti, tn \in c \textit{ tal que } i < n \textit{ e } Suc_w(\imath, ti, tn) = e \\ & \quad \textit{ou } Fal_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V_t^w(Group(gr)(e)) = \begin{cases} 1 & \textit{, se existe } tn \in c \textit{ tal que } tn > t \textit{ e } V_{tn}^w(Execute_{gr}(e)) = 1 \\ 0 & \textit{, caso contrário} \end{cases}$$

Observação 7.3.1

$c(1)$ *indica o próximo mundo de cada caminho originado do mundo atual.*

$CAMINHOS(w_t) \subseteq CAMINHOS(w)$ *é o conjunto de caminhos de* w *a partir do estado* t*, seja* $w_t \in w$.

O operador $\mathcal{W}$ *é definido como um operador derivado de* $\mathcal{U}$ *e da t-conorma* S*:*

$$\varphi\,\mathcal{W}\,\phi \stackrel{\text{def}}{=} \mathrm{S}(\varphi\mathcal{U}\phi), \Box\varphi) \tag{7.8}$$

A partir de então, é possível determinar a satisfazibilidade ou não das fórmulas BDI fuzzy versão I.

Definição 7.3.4 *Seja um modelo* M*, uma fórmula* $\varphi \in L^{MF}$ *e* k *uma constante veritativa dada pelo sistema, onde* $k \in [0,1]$ *e os graus de verdade maiores ou iguais a ela são considerados verdadeiros.* φ *é satisfazível para um modelo* M *dada uma situação* (w,t) *do modelo* M *tal que* $t \in T_w$ *e* $w \in W$*, sse* $V_t^w(\varphi) \geq k$*, denotado por* $M, V, w_t \models^{k_I} \varphi$*; e* φ *é insatisfazível em* M*, denotado por* $M \not\models^{k_I} \varphi$*, se não existe uma situação deste modelo* M *tal que* $V_t^w(\varphi) \geq k$*, ou em outras palavras:* $V_t^w(\varphi) < k$ *para qualquer elemento de* S_w.

φ é verdadeira em um modelo M, denotado por $M, V \models^{k_I} \varphi$, sse existe pelo menos uma situação (w, t) em cada mundo possível w de M tal que $V_t^w(\varphi) \geq k$, ou seja, φ é satisfeita em todos os mundos w de M; e φ é contingente se existem mundos w e w' de M tais que $V_t^w(\varphi) \geq k$ e $V_t^{w'}(\varphi) < k$, isto é, φ é contigente se $M, V, w\_ \models^{k_I} \varphi$ e $M, V, w'\_ \not\models^{k_I} \varphi$; e φ é falsa sse ela é insatisfazível para qualquer modelo M. Por fim, φ é válida, denotado por $\models^{3M} \varphi$ sse para qualquer instante, mundo e modelo $V(\varphi) \geq k$ – o conjunto das fórmulas válidas de $\mathcal{L}^{MF}$ é definido por $FVMF = \{\varphi \in L^{MF} \ / \ \models^{k_I} \varphi\}$ –; e é insatisfazível quando não existe um modelo para a mesma, isto é, para qualquer $\langle M, w, t\rangle$, φ é falsa, denotado por $\not\models^{k_I} \varphi$.

Note que os mundos fuzzy não podem mais ser pensados como um conjunto de sentenças verdadeiras tal como acontece nas lógicas modais clássicas. Pois, nela haverão sentenças as quais podem ser consideradas verdadeiras ou falsas. Em função disso, a caracterização das crenças, desejos e intenções não é tão semelhante à descrita na Log^{3M}. Isto é, seja $\overline{\mathcal{M}}$ um conectivo modal e $R'_{\overline{\mathcal{M}}}$ a relação de acessibilidade correspondente a tal conectivo; então, se $M, V, w_t \models^{k_I} \overline{\mathcal{M}}(\varphi)$, então, $V_t^w(\overline{\mathcal{M}}(\varphi)) \geq k$ e todo $w_t R'_{\overline{\mathcal{M}}} w'$ tal que $M, V, w_t \models^{k_I} \varphi$, isto é, $V_t^{w'}(\varphi) \geq k$. E se existe w' tal que $w_t R'_{\overline{\mathcal{M}}} w'$ e $V_t^{w'}(\varphi) < k$, então, $M, V, w_t \not\models^{k_I} \overline{\mathcal{M}}(\varphi)$, ou seja: $V_t^w(\overline{\mathcal{M}}(\varphi)) < k$.

Da mesma forma que há a definição de uma fórmula satisfazível, define-se a satisfazibilidade de um conjunto de fórmulas Γ de L^{MF} como sendo um modelo M e uma situação (w, t) à qual todas as fórmulas $\varphi \in \Gamma$ são verdadeiras, ou seja, $V_t^w(\varphi) \geq k$. Dada as definições anteriores, pode-se definir o conceito de consequência semântica e lógica BDI fuzzy versão I.

Definição 7.3.5 *Seja a linguagem* BDI *fuzzy $\mathcal{L}^{MF}$ e $\Gamma \subseteq L^{MF}$. φ é uma consequência semântica de Γ, denotado por $\Gamma \models^{k_I} \varphi$ se todas as situações em que Γ for satisfazível φ também é, ou seja, $\forall(w, t)$ / se $M, w, t \models^{k_I} \Gamma$ então $M, w, t \models^{k_I} \varphi$.*

Mantém-se para esta lógica a mesma definição 5.2.1 de Log^{3M}, para que o agente acredite em suas intenções e desejos, e deseje suas intenções.

Por fim, a lógica tri-modal BDI fuzzy versão I é definida.

Definição 7.3.6 *Seja a linguagem modal $\mathcal{L}^{MF}$, a lógica* BDI *fuzzy versão I definida sobre esta linguagem é a lógica Log^{FM_I} onde*

$$Log^{FM_I} = \langle L^{MF}, \models^{k_I}\rangle$$

7.3.2 Semântica da Lógica BDI Fuzzy versão II

A sintaxe da versão II é a mesma da versão I. No entanto, na versão II da lógica BDI fuzzy, as relações R e R' possuem graus de verdade dentro do intervalo [0,1]. O sentido de R' ser valorada, além de ser uma outra forma de se obter o valor das sentenças que envolvem os operadores modais Bel, Des e Int; também mudará a caracterização das atitudes mentais, uma vez que tais relações não levarão apenas a mundos em que as crenças, os desejos e as intenções do agentes são considerados verdadeiros (esta caracterização será detalhada posteriormente). A consideração de um valor de verdade fuzzy para R também implicará numa diferença significativa da semântica, tanto na interpretação dos conectivos temporais, quanto na interpretação dos conectivos de ações.

Para valorar R definir-se-á uma nova função de valoração υ. Para que as relações de acessibilidade possuam valores verdade definir-se-ão as mesmas como funções compostas.

Definição 7.3.7 *Sejam* $\begin{array}{l} B : L_{Gr} \rightarrow \wp\{W \times T \times W\} \rightarrow [0,1] \\ D : L_{Gr} \rightarrow \wp\{W \times T \times W\} \rightarrow [0,1] \\ I : L_{Gr} \rightarrow \wp\{W \times T \times W\} \rightarrow [0,1] \end{array}$ *onde B é uma relação de acessibilidade de crença, D é uma relação de acessibilidade de desejo e I é uma relação de acessibilidade de intenção. $\langle B, D, I\rangle$ são relações de acessibilidade* BDI *do tipo II se satisfazem as seguintes condições:*

- $\exists\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(B(\imath)) \wedge \langle w,t,w''\rangle \in dom(D(\imath)) \wedge w'' \sqsubseteq w')$*;*
- $\forall\langle w,t,w'\rangle.(\langle w,t,w'\rangle \in dom(D(\imath)) \rightarrow \exists\langle w,t,w''\rangle.(\langle w,t,w''\rangle \in dom(I(\imath)) \wedge w'' \sqsubseteq w')$*; e*
- $\exists\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(B(\imath)) \wedge \langle w,t,w''\rangle \in dom(I(\imath)) \wedge w'' \sqsubseteq w')$.

Note que a interpretação das condições de satisfação é a mesma da lógica BDI explicitada no capítulo anterior. Isto é feito para que as inter-relações entre as atitudes mentais sejam as mesmas descritas por Bratman em sua Tese de Assimetria (já referenciada na seção 5.1 desta dissertação), bem como, em seu modelo de desejo-crença ([Bra87]).

Definição 7.3.8 *Um modelo M nesta lógica é uma tupla $\langle W, T, \boldsymbol{D}, R, B, D, I, \Phi, E\rangle$ onde W, T, D, Φ, E são os mesmos da definição 7.3.1. $\langle B, D, I\rangle$ são relações de acessibilidade* BDI *do tipo II (definição 7.3.7). E R é uma função que mapeia um arco segundo um dado agente ou grupo de agentes a um valor do intervalo* $[0, 1]$*, ou seja* $R : L^{Gr} \times T \times T \rightarrow [0, 1]$.

É possível definir dois sentidos para a função R. O primeiro, o qual será usado neste trabalho, captura a noção de eventos já que os operadores $Sucesso$, $Falha$ e $Executado$ são definidos sobre ela. O segundo captura a noção de proximidade entre mundos possíveis, com o valor 1 correspondendo à identidade de mundos possíveis e um valor 0 indicando que os valores das proposições verdadeiras em um estado t não fornecem qualquer indicação de verdade no outro estado t' tal que tRt'. Para isso, tal função deve ser simétrica e satisfazer uma forma de "transitividade fraca" expresso como: $R(t,t'') \geqslant T(R(t,t'), R(t',t''))$. As relações R, quando possuem o segundo sentido, elas são chamadas de relações de similaridades. Tal idéia foi apresentada por Hájek em [HH96, Háj98] e por Godo e Rodriguez em [GR02, RGG96]. Em [GR02] é aplicada, pelos autores, para revisar e atualizar bases de conhecimento.

Há ainda duas funções de valoração: para as fórmulas (V) e para as relações entre estados de tempo (υ).

Definição 7.3.9 *A atribuição dos valores verdade das variáveis (livres) são atribuídos em cada situação (w, t) por ρ. ρ é definido similarmente à sua definição na semântica da lógica* BDI *fuzzy versão I, bem como as definições de $w_t(f)$, $w_t(P)$, $w_t(x)$ e $w_t(a)$.*

V é a função de valoração que mapeia uma fbf de uma situação (w,t) em um valor do intervalo [0,1], ou seja: $V : S_w \times L^{MF} \rightarrow [0,1]$ e a função υ de valoração de relação de tempos é definida como $\upsilon : T \times T \times Ling(\mathcal{L}^E) \rightarrow [0,1]$ e captura a noção de quão realizado foi o evento $e \in Ling(\mathcal{L}^e)$.

Além dos símbolos de relações serem interpretados como relações fuzzy, os conectivos proposicionais, símbolos modais e temporais, inclusive $Success$, $Fail$ e $Execute$ possuírem uma semântica fuzzy. As relações de acessibilidade entre mundos e entre instantes de tempo também tem valoração fuzzy. Portanto, a definição dos conectivos da linguagem BDI *fuzzy versão II, a partir das funções V e de υ, segue a interpretação abaixo:*

$V_t^w(R(t1,\ldots,tn)) = \mu_R(V_t^w(\tau_1),\ldots,V_t^w(\tau_n))$.

$V_t^w(\neg\varphi) = \mathrm{N}(V_w(\varphi))$

$V_t^w((\varphi \wedge \phi)) = T(V_w^t(\varphi), V_t^w(\phi))$,

$V_t^w((\varphi \vee \phi)) = S(V_t^w(\varphi), V_w^t(\beta))$

$V_t^w((\varphi \rightarrow \phi)) = \mathrm{J}(V_w(\varphi), V_w(\phi))$,

$V_t^w((\varphi \leftrightarrow \phi)) = \mathrm{B}(V_w^t(\varphi), V_w^t(\phi))$,

$V_t^w(\forall x.(\phi)) = inf\{V_w^t(\phi) \,/\, V'{}_t^w(y) = V_t^w(y)\,\forall y.(y \neq x)\}$

$V_t^w(\exists x.(\phi)) = sup\{V_w^t(\phi) \,/\, V'{}_t^w(y) = V_t^w(y)\,\forall y.(y \neq x)\}$

$V_t^w(\Diamond\varphi) = sup\{\mathrm{J}(\upsilon(t,t',e),\, V_{t'}^w(\varphi))\,/\,tRt'\})$,

$V_t^w(\Box\varphi) = inf\{\mathrm{J}(\upsilon(t,t',e),\, V_t^{w'}(\varphi))\,/\,tRt'\}$,

$V_t^w(\bigcirc\varphi) = inf\{\mathrm{J}(\upsilon(t,t',e),\, V_{t'}^w(\varphi))\,/\,tRt' \textit{ e } t' = c(1)\}$

$V_t^w(\varphi\,\mathcal{U}\,\phi) = min\{inf\{V_j^w(\varphi)\,/\; j > t \textit{ e } j,t \in c\},\, V_{t'}^w(\phi)\}$, *onde*
$t' = max\{j \in c\,/\; V_j^w(\neg\phi) \geq k\} + 1$

$V_t^w(Inev(\varphi)) = inf\{\mathrm{J}(\upsilon(t,t',e),\, V_{t'}^w(\varphi))\,/\,t' \in CAMINHO(w_t)\}$

$V_t^w(Opt(\varphi)) = sup\{\mathrm{J}(\upsilon(t,t',e),\, V_{t'}^w(\varphi))\,/\,t' \in CAMINHO(w_t)\}$

$V_t^w(Bel_\imath\varphi) = inf\{\mathrm{J}(B(w,w'),\, V_t^{w'}(\varphi))\,/\,wBw'\}$

$V_t^w(Des_\imath\varphi) = inf\{\mathrm{J}(D(w,w'),\, V_t^{w'}(\varphi))\,/\,wDw'\}$

$V_t^w(Int_\imath\varphi) = inf\{\mathrm{J}(I(w,w'),\, V_t^{w'}(\varphi))\,/\,wIw'\}$

$$V_{tn}^w(Success_\imath(e)) = \begin{cases} \upsilon(ti,tn,e) & \textit{, se existe } ti,tn \in c \textit{ tal que } i < n \\ & \textit{ e } Suc_w(\imath,ti,tn) = e \\ \mathrm{N}(\upsilon(ti,tn,e)) & \textit{, se } Fal_w(\imath,ti,tn) = e \end{cases}$$

$$V^w_{tn}(Fail_\imath(e)) = \begin{cases} \upsilon(ti, tn, e) & \textit{, se existe } ti, tn \in c \textit{ tal que } i < n \\ & \textit{e } Fal_w(\imath, ti, tn) = e \\ \mathrm{N}(\upsilon(ti, tn, e)) & \textit{, se } Suc_w(\imath, ti, tn) = e \end{cases}$$

$$V^w_{tn}(Execute_\imath(e)) = \begin{cases} \upsilon(ti, tn, e) & \textit{, se existe } ti, tn \in c \textit{ tal que } i < n \\ & \textit{e } Suc_w(\imath, ti, tn) = e \\ & \textit{ou } Fal_w(\imath, ti, tn) = e \\ 0 & \textit{, caso contrário} \end{cases}$$

$$V^w_t(Group(gr)(e)) = \begin{cases} 1 & \textit{, se existe } tn \in c \textit{ tal que } tn > t \\ & \textit{e } V^w_{tn}(Execute_{gr}(e)) = \upsilon(ti, tn, e) \\ 0 & \textit{, caso contrário} \end{cases}$$

A partir daí, é possível determinar a satisfazibilidade das fórmulas BDI fuzzy, segundo a semântica da versão II semelhantemente à definição 7.3.4.

Definição 7.3.10 *Seja um modelo $M = \langle W, T, D, R, R', \Phi\rangle$, uma fórmula $\varphi \in L_{FM_{II}}$ e k uma constante veritativa dada pelo sistema, onde $k \in [0, 1]$ e os graus de verdade maiores ou iguais a ela são considerados verdadeiros. φ é satisfazível para um modelo M se existe uma situação (w, t) do modelo M tal que $t \in T_w$ e $w \in W$, $V^w_t(\varphi) \geq k$, denotado por, $M, V, \upsilon, w_t \models^{K2} \varphi$. E φ é dita insatisfazível para M, denotado por $M \not\models^{k_{II}} \varphi$, se não existe uma situação deste modelo tal que $V^w_t(\varphi) \geq k$, ou em outras palavras: $V^w_t(\varphi) < k$ para qualquer elemento de S_w. Uma fórmula φ é verdadeira em um modelo M, denotado por $M, V, \upsilon \models \varphi$, se existe pelo menos uma situação (w, t) em cada mundo possível w de M tal que $V^w_t(\varphi) \geq k$, ou seja, φ é satisfeita em todos os mundos w de M; φ é falsa em M sse ela é insatisfazível para M; e φ é contingente se existem mundos w e w' de M tais que $V^w_t(\varphi) \geq k$ e $V^{w'}_t(\varphi) < k$, isto é, φ é contingente sse $M, V, \upsilon, w\_ \models^{k_{II}} \varphi$ e $M, V, \upsilon, w'\_ \not\models^{k_{II}} \varphi$. Por fim, φ é dita universalmente válida se ela for verdadeira em todos os modelos, denotado por $\models^{k_{II}} \varphi$. Podendo ainda ser validada em um dada classe de modelos, denotado por $C \models^{k_{II}} \varphi$.*

O conceito de satisfazibilidade para um conjunto de fórmulas $\Gamma \subseteq L^{MF}$ é o mesmo da lógica BDI fuzzy versão I, explicitado na seção anterior. Dada as definições anteriores, pode-se definir o conceito de consequência semântica.

Definição 7.3.11 *Seja a linguagem modal fuzzy $\mathcal{L}^{MF}$ e $\Gamma \subseteq L^{MF}$. $\varphi \in L^{MF}$ é uma consequência semântica de Γ, denotado por $\Gamma \models^{k_{II}} \varphi$ se $\forall(w, t)$ / se M, V,*
$\upsilon, w_t \models^{k_{II}} \Gamma$ então $M, w, t \models^{k_{II}} \varphi$, ou seja, se todos as situações em que Γ for satisfazível φ também o é.

Definição 7.3.12 *A partir de um dado mundo w, todo mundo de crença – originados deste w – se relaciona com cada mundo de intenções e de desejos – também provenientes deste w – via relação I e D, respectivamente, onde seja b, d e i mundos de crenças, desejos e intenções (respectivamente) provenientes de uma mesma situação (w, t), $I(b, i) \geq k$ e $D(b, d) \geq k$. E todo mundo de desejo se relaciona com cada mundo de intenções via relação I onde $I(d, i) \geq k$.*

A definição 7.3.12 é necessária pelo mesmo motivo da definição 5.2.1 com relação às lógicas BDI e BDI fuzzy versão I. Já que o agente ou grupo de agentes deve acreditar que tem uma intenção e um dado desejo assim que ele tiverem tal intenção ou desejo; devendo também desejar sua intenção quando tiver alguma.

A lógica BDI fuzzy versão II é definida a seguir.

Definição 7.3.13 *Seja a linguagem modal $\mathcal{L}^{MF}$, a lógica* BDI *fuzzy versão II definida sobre esta linguagem é a lógica $Log^{FM_{II}}$ onde*

$$Log^{FM_{II}} = \langle L^{MF}, \models^{k_{II}} \rangle.$$

Caracterização de Crenças, Desejos e Intenções para a Semântica da Lógica BDI Fuzzy Versão II

A caracterização das crenças, desejos e intenções para a semântica da versão II é um pouco diferente da caracterização das mesmas na versão I da lógica BDI fuzzy. Na versão II, como foi dito anteriormente, B, D e I não relacionam dois mundos em cujo segundo deles estão a satisfação das sub-fórmulas pertencentes ao escopo de Bel, Des e Int (respectivamente).

É interessante, portanto, a princípio, entender o que significa a valoração de uma relação de acessibilidade. Primeiramente, nota-se que R' funciona similarmente à R na interpretação de $\Box$. Na lógica modal clássica, a relação entre os mundos (w e w' por exemplo) liga mundos que são compatíveis. No caso de $\Box$, esta compatibilidade significa que a sub-fórmula dentro do escopo de $\Box$ é verdadeira no mundo w' tal que wRw'. Na versão II da lógica BDI fuzzy, haverá um grau de compatibilidade, no qual, ele ser considerado verdadeiro é suficiente para que a sub-fórmula φ, seja considerada verdadeira, em w'.

Dessa forma, seja $R'_{\overline{\mathcal{M}}}$ uma relação de acessibilidade correspondente ao operador modal $\overline{\mathcal{M}}$, então, a primeira é definida como se segue: $R'_{\overline{\mathcal{M}}} : L^{Gr} \to \wp\{W \times T \times W\} \to [0, 1]$. O que quer dizer que $\overline{\mathcal{M}}$ possui um valor verdade e mapeia um dado tempo t de um mundo possível w em outro mundo w'; tal valor é considerado verdadeiro se em w' e no tempo t, a sub-fórmula é considerada verdadeira ou $R'_{\overline{\mathcal{M}}}$ é considerada falsa, pois

$$V^w_t(\overline{\mathcal{M}}_\imath(\varphi)) = inf\{\mathrm{J}(R'_{\overline{\mathcal{M}}}(w, w'),\, V^t_{w'}(\varphi)) \,/\, wR'_{\overline{\mathcal{M}}}w'\}$$

.

Devido à mudança na definição das relações de acessibilidade e o que provocou a mudança na caracterização das atitudes (explicitada anteriormente). Há de se redefinir as relações entre as atitudes mentais; a seguir, definir-se-á as relações entre atitudes mentais segundo a semântica desta lógica BDI fuzzy versão II.

Definição 7.3.14 *Seja $\mathcal{M}$ e $\overline{\mathcal{M}}$ dois dos operadores modais Bel, Des e Int, tal que $\mathcal{M} \neq \overline{\mathcal{M}}$, definidos pelas relações de acessibilidade R' e $\overline{R'}$, respectivamente.*

$R'(\imath) = \overline{R'}(\imath)$ *sse* $R'(\imath) \subseteq \overline{R'}(\imath)$ *e* $\overline{R'}(\imath) \subseteq R'(\imath)$

$dom(R'^{\,w}_{\,t}(\imath)) \cap_{sub} dom(\overline{R'}^{w}_{t}(\imath)) = \{\langle w, t, w' \rangle \,/\, \langle w, t, w' \rangle \in dom(\overline{R'}(\imath))$ *e* $\exists \langle w, t, w'' \rangle \in dom(R'(\imath))$ *t.q.* $w'' \sqsubseteq w'$

$dom(R'^{\,w}_{\,t}(\imath)) \cap_{sup} dom(\overline{R'}^{w}_{t}(\imath)) = \{\langle w,t,w'\rangle \,/\, \langle w,t,w'\rangle \in dom(\overline{R'}(\imath))$ *e* $\exists\langle w,t,w''\rangle \in dom(R'(\imath))$ *t.q.* $w' \sqsubseteq w''$

$dom(R'^{\,w}_{\,t}(\imath)) \subseteq dom(\overline{R'}^{w}_{t}(\imath))$ *sse* $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(R'(\imath) \rightarrow (\langle w,t,w''\rangle \in dom(\overline{R'}(\imath)) \wedge CAMINHOS(w') \subseteq CAMINHOS(w''))$

$dom(R'^{\,w}_{\,t}(\imath)) \subseteq_{sub} dom(\overline{R'}^{w}_{t}(\imath))$ *sse* $\forall\langle w,t,w'\rangle.(\langle w,t,w'\rangle \in dom(R'(\imath)) \rightarrow \exists\langle w,t,w''\rangle.(\langle w,t,w''\rangle \in dom(\overline{R'}(\imath))) \wedge w'' \sqsubseteq w'')$

$dom(R'^{\,w}_{\,t}(\imath)) \subseteq_{sup} dom(\overline{R'}^{w}_{t}(\imath))$ *sse* $\forall\langle w,t,w'\rangle.(\langle w,t,w'\rangle \in dom(R'(\imath)) \rightarrow \exists\langle w,t,w''\rangle.(\langle w,t,w''\rangle \in dom(\overline{R'}(\imath)) \wedge w'' \sqsubseteq w'))$

A partir desta definição, as demonstrações de que a lógica BDI fuzzy versão II também obedece aos aspectos do modelo desejo-crença e à tese de assimentria propostos por Bratman [Bra87] são facilitadas. Estes aspectos serão abordados na próxima seção.

7.3.3 As Lógicas BDI Fuzzy como Extensões da Formalização da Tese de Assimetria

Ver-se-á a seguir as demonstrações de que a inter-relação entre crenças e intenções sugeridas por Bratman na sua tese de assimetria é obtida pelas lógicas BDI fuzzy.

O conjunto de provas a ser explicitado, em seguida, segue a mesma sequência e tem a mesma finalidade do conjunto de prova exposto no capítulo anterior cuja finalidade era demonstrar que a nova lógica BDI proposta pelo autor deste trabalho é uma lógica que atende aos requisitos do modelo desejo-crença e tese de assimetria definidos por Bratman em [Bra87].

Proposição 7.3.1 $\models^{k_I} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Pela definição das relações de acessibilidade em 7.3.2, tem-se que $D^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$, então $\exists w', w''.(w' \in B^w_t(\imath) \wedge w'' \in D^w_t(\imath) \wedge w'' \sqsubseteq w')$, ou seja, existe pelo menos um mundo de crença e de desejo, os quais, o segundo é submundo do primeiro. Agora assuma que $\models^{k_I} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Daí, para algum $\langle M,V,w,t\rangle$ arbitrário, $M,V,w_t \models^{k_I} Des_\imath(\varphi)$ e $M,V,w_t \models^{k_I} Bel_\imath(\neg Opt(\varphi))$. A partir da interpretação do conectivo modal Des, conclui-se que para todo $x' \in D^w_t(\imath)$, $V^{x'}_t(Opt(\varphi)) \geq k$; e pela semântica de Bel, conclui-se que para todo $y'' \in B^w_t(\imath)$, $V^{y''}_t(\neg Opt(\varphi)) \geq k$. Por consequência de $D^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$, existe pelo menos um $w' \in D^w_t(\imath)$ tal que $V^{w'}_t(\neg Opt(\varphi)) > k$ ou existe um $w'' \in B^w_t(\imath)$ tal que $V^{w''}_t(Opt(\varphi)) \geq k$. Portanto, em ambos os casos (considerando a interpretação dada ao conectivo $\neg$) há uma contradição. Logo, a suposição há de ser falsa. ■

Corolário 7.3.1 $D^w_t(\imath) \cap_{sub} B^w_t(\imath) \neq \emptyset$ *sse* $\models^{k_I} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Proposição 7.3.2 $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$.

PROVA: Pela definição 7.3.2, tem-se que $D(\imath) \subseteq_{sup} I(\imath)$, então, $\forall x'.(x' \in D_t^w(\imath) \rightarrow \exists w''.(w'' \in I_t^w(\imath) \wedge w'' \sqsubseteq x')$, ou seja, para todo $x' \in D(\imath)$ existe um $w'' \in I(\imath)$ em que $w'' \sqsubseteq x'$. Agora assuma que $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$ é falso. Portanto, $M, V, w_t \models^{k_I} Int_\imath(Opt(\varphi))$ e $M, V, w_t \models^{k_I} \neg Des_\imath(Opt(\varphi))$, para um dado $\langle M, V, w, t\rangle$ arbitrário. A partir da interpretação dos conectivos Int e Des e dada uma constante veritativa k, sabe-se, respectivamente, que: para todo $y'' \in I_t^w(\imath)$, $V_t^{y''}(Opt(\varphi)) \geq k$ e existe um $w' \in D_t^w(\imath)$ tal que $V_t^{w'}(\neg Opt(\varphi))$. Mas, como $D(\imath) \subseteq_{sup} I(\imath)$ e existe um $w' \in D_t^w(\imath)$ tal que $V_t^{w'}(\neg Opt(\varphi))$, então, existe um $w'' \in I_t^w(\imath)$ tal que $V_t^{w''}(\neg Opt(\varphi)) \geq k$. Considerando a interpretação de $\neg$, isto é uma contradição, portanto, a suposição é falsa. ■

Corolário 7.3.2 $D(\imath) \subseteq_{sup} I(\imath)$ *sse,* $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$.

Proposição 7.3.3 $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Pela definição 7.3.2, tem-se que $I(\imath) \cap_{sub} B(\imath) \neq \emptyset$, então, $\exists w', w''.(w'' \in B_t^w(\imath) \wedge w' \in I_t^w(\imath) \wedge w' \sqsubseteq w'')$, ou seja, existe algum $w'' \in B_t^w(\imath))$ e um $w' \in I_t^w(\imath))$ tal que $w' \sqsubseteq w''$. Agora suponha que $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Daí, para algum $\langle M, V, w, t\rangle$ qualquer,$M, V, w_t \models^{k_I} Int_\imath(\varphi)$ e $M, V, w_t \models^{k_I} Bel_\imath(\neg Opt(\varphi))$. Dada uma constante veritativa k; pela semântica de Int, pode-se concluir que para todo $x' \in I_t^w(\imath)$, $V_t^{x'}(Opt(\varphi)) \geq k$. Todavia, pela semântica de Bel, pode-se concluir que para todo $y'' \in B_t^w(\imath)$, $V_t^{y''}(\neg Opt(\varphi)) \geq k$. Por consequência de $I(\imath) \cap_{sub} B(\imath) \neq \emptyset$, existe pelo menos um $w' \in D(\imath)$ tal que $V_t^{w'}(\neg Opt(\varphi)) \geq k$. Portanto, considerando a interpretação de $\neg$, há uma contradição. Logo, a suposição há de ser falsa. ■

Corolário 7.3.3 $I_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$ *sse* $\models^{k_I} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

Teorema 7.3.1 *Se* $Int_\imath(Opt(\varphi))$*, então,* $\neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Prova trivial por silogismo hipotético, a partir dos corolários 7.3.1 e 7.3.2. ■

Teorema 7.3.2 $I_t^w(\imath) \cap_{sub} B_t^w(\imath) \neq \emptyset$

PROVA: Prova trivial a partir do teorema 7.3.1 e da proposição 7.3.3. ■

A noção dada pela tese de assimetria de Bratman também é inerente à lógica BDI fuzzy versão II.

Proposição 7.3.4 $\models^{k_{II}} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.

PROVA: Pela definição de das relações de acessibilidade, tem-se que $dom(D_t^w(\imath)) \cap_{sub} dom(B_t^w(\imath)) \neq \emptyset$, então, $\exists\langle w, t, w'\rangle, \langle w, t, w''\rangle.(\langle w, t, w'\rangle \in dom(B(\imath)) \wedge \langle w, t, w''\rangle \in dom(D(\imath)) \wedge w'' \sqsubseteq w')$. Agora assuma que $\models^{k_{II}} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Daí, para algum $\langle M, V, w, t\rangle$ arbitrário, $M, V, w_t \models^{k_{II}} Des_\imath(\varphi)$ e $M, V, w_t \models^{k_{II}} Bel_\imath(\neg Opt(\varphi))$. A partir da interpretação do conectivo modal Des, conclui-se que para todo $x' \in dom(D_t^w(\imath))$, $V_t^{x'}(Opt(\varphi)) \geq k$ independente do valor verdade de $D(w, x')$; e pela semântica de Bel, conclui-se que para todo $y'' \in dom(B_t^w(\imath))$, $V_t^{y''}(\neg Opt(\varphi)) \geq k$. Por consequência de $dom(D_t^w(\imath)) \cap_{sub} dom(B_t^w(\imath)) \neq \emptyset$, existe pelo menos um $w' \in D_t^w(\imath)$ tal que $V_t^{w'}(\neg Opt(\varphi)) > k$ ou existe um $w'' \in B_t^w(\imath)$ tal que $V_t^{w''}(Opt(\varphi)) \geq k$. Portanto, em ambos os casos (considerando a interpretação dada ao conectivo $\neg$) há uma contradição. Logo, a suposição há de ser falsa. ■

Corolário 7.3.4 *$dom(D_t^w(\imath)) \cap_{sub} dom(B_t^w(\imath)) \neq \emptyset$ sse $\models^{k_{II}} Des_\imath(\varphi) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.*

Proposição 7.3.5 *$\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$.*

PROVA: Pela definição 7.3.7, tem-se que $dom(D(\imath)) \subseteq_{sup} dom(I(\imath))$. Logo, $\forall \langle w, t, x' \rangle . (\langle w, t, x' \rangle \in dom(D(\imath)) \rightarrow \exists \langle w, t, w'' \rangle . (\langle w, t, w'' \rangle \in dom(I(\imath)) \wedge w'' \sqsubseteq x')$ ou seja, para todo $\langle w, t, x' \rangle \in dom(D(\imath))$ existe algum $\langle w, t, w'' \rangle \in dom(I(\imath))$ em que $w'' \sqsubseteq x'$. Agora assuma que não é o caso de $\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$, portanto, $M, V, w_t \models^{k_{II}} Int_\imath(Opt(\varphi))$ e $M, V, w_t \models^{k_{II}} \neg Des_\imath(Opt(\varphi))$, para um dado $\langle M, V, w, t \rangle$ arbitrário. A partir da interpretação dos conectivos Int e Des e dada uma constante veritativa k, sabe-se, respectivamente, que: para todo $y'' \in dom(I_t^w(\imath))$, $V_t^{y''}(Opt(\varphi)) \geq k$ e existe um $w' \in dom(D_t^w(\imath))$ tal que $V_t^{w'}(\neg Opt(\varphi))$. Mas, como $dom(D(\imath)) \subseteq_{sup} dom(I(\imath))$ e existe um $w' \in dom(D_t^w(\imath))$ tal que $V_t^{w'}(\neg Opt(\varphi))$, então existe um $\langle w, t, w'' \rangle \in dom(I(\imath))$ tal que $V_t^{w''}(\neg Opt(\varphi)) \geq k$ independente do valor verdade de $I(w, w'')$. Considerando a interpretação de $\neg$, isto é uma contradição, portanto, a suposição é falsa. ■

Corolário 7.3.5 *$dom(D(\imath)) \subseteq_{sup} dom(I(\imath))$ sse, $\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$.*

Proposição 7.3.6 *$\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.*

PROVA: Pela definição 7.3.7, tem-se que $dom(I(\imath)) \cap_{sub} dom(B(\imath)) \neq \emptyset$. Logo, existe algum $\langle w, t, w'' \rangle \in B(\imath))$ e $\langle w, t, w' \rangle \in dom(I(\imath))$ tal que $w' \sqsubseteq w''$. Agora suponha que $\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$ é falso. Daí, para algum $\langle M, V, w, t \rangle$ qualquer,$M, V, w_t \models^{k_{II}} Int_\imath(\varphi)$ e $M, V, w_t \models^{k_{II}} Bel_\imath(\neg Opt(\varphi))$. Dada uma constante veritativa k; pela semântica de Int, pode-se concluir que para todo $x' \in dom(I_t^w(\imath))$, $V_t^{x'}(Opt(\varphi)) \geq k$ independente do valor verdade de $I(w, x')$. Todavia, pela semântica de Bel, pode-se concluir que para todo $\langle w, t, y'' \rangle \in dom(B(\imath))$, $V_t^{y''}(\neg Opt(\varphi)) \geq k$ independente do valor verdade de $B(w, y'')$. Por consequência de $dom(I(\imath)) \cap_{sub} dom(B(\imath)) \neq \emptyset$, existe pelo menos um $\langle w, t, w' \rangle \in dom(D(\imath))$ tal que $V_t^{w'}(\neg Opt(\varphi)) \geq k$. Portanto, considerando a interpretação de $\neg$, há uma contradição. Logo, a suposição há de ser falsa. ■

Corolário 7.3.6 *$dom(I_t^w(\imath)) \cap_{sub} dom(B_t^w(\imath)) \neq \emptyset$ sse $\models^{k_{II}} Int_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$.*

Teorema 7.3.3 *Se $Int_\imath(Opt(\varphi))$ então $\neg Bel_\imath(\neg Opt(\varphi))$.*

PROVA: Prova trivial por silogismo hipotético, a partir das proposições 7.3.4 e 7.3.5. ■

Teorema 7.3.4 *$dom(I_t^w(\imath)) \cap_{sub} dom(B_t^w(\imath)) \neq \emptyset$*

PROVA: Prova trivial a partir do teorema 7.3.3 e da proposição 7.3.6. ■

7.4 Teoria Formal BDI Fuzzy

Apresentar-se-á nesta seção uma teoria formal cabível às duas lógicas BDI versão I e II. Ela é denotada por T^{MF} e será definida semelhantemente à T^{3M}. São mantidos os mesmos axiomas para as crenças desejos e intenções. Quanto aos axiomas referentes às inter-relações entre as atitudes mentais, eles também permanecem os mesmos, pois elas continuam sendo fundamentados sobre a mesma teoria. E as regras de inferência também são mantidas.

Logo, $T^{MF} = \langle L^{MF'}, \{, K_B, D_B, 4_B, 5_B\, K_D, D_D,\, K_I, D_I\} \cup \{A1, \ldots, A5\}, \{MP, Gen, Nes, Nes_B, Nes_D, Nes_I\}\rangle$ tal que $L^{MF'} = \langle \Sigma^{MF'}, \mathcal{G}^{MF'}\rangle$ é uma linguagem BDI fuzzy simplificada, onde

$\Sigma^{MF'} = \{X \cup \Sigma_d \cup \Sigma_{L_{FM'}} \cup \Sigma_P \cup Ling(\mathcal{L}^E) \cup Ling\mathcal{L}^{Gr}\}$ sendo

$\Sigma_{L_{FM'}} = \Sigma_{L^E} \cup \{\neg, \rightarrow, \forall, \bigcirc, \mathcal{U},\, Inev,\, Bel,\, Des,\, Int,\, Fail\}$; e

$\mathcal{G}^{FM'} = \mathcal{G}^{MF} - \{F_3, F_4, F_6, F_7, F_9, F_{13}, F_{18}, F_{24}, F_{26}, F_{27}\}$.

O conjunto de axiomas K, D ,4 ,5 aplicados à Bel $(K_B, D_B, 4_B, 5_B)$ e os axiomas K e D da lógica modal aplicados à Des e Int (K_D, D_D) e (K_I, D_I).

O conjunto de axiomas $\{A1, \ldots, A5\}$ são os axiomas mostrados no capítulo anterior e pode também ser acrescidos do(s) axioma(s) $\{A6_a, A6_b, A6_c\}$ referente(s) as estratégias de compromisso cego, simples e mente aberta.

E $\{MP, Gen, Nes, Nes_B, Nes_D, Nes_I\}$ são as regras de inferência de Modus Ponens, Generalização (do conectivos $\forall$), Nescessitação (dos conectivos $\Box$, Bel, Des e Int).

Observação 7.4.1 *Os teoremas envolvendo operadores não definidos pelas regras gramaticais explicitadas em* $\mathcal{G}^{MF'}$ *podem ser provados usando as mesmas equivalências explicitadas na subseção 5.3.*

Assim, os axiomas e suas respectivas condições semânticas são explicitadas a seguir:

[A1] $Des_\imath(Opt(\varphi)) \rightarrow \neg Bel_\imath(\neg Opt(\varphi))$

[SC1] $\exists\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(B(\imath)) \wedge \langle w,t,w''\rangle \in dom(D(\imath)) \wedge w'' \sqsubseteq w')$

[A2] $Int_\imath(Opt(\varphi)) \rightarrow Des_\imath(Opt(\varphi))$

[SC2] $\forall\langle w,t,w'\rangle.(\langle w,t,w'\rangle \in dom(D(\imath)) \rightarrow \exists\langle w,t,w''\rangle.(\langle w,t,w''\rangle \in dom(I(\imath)) \wedge w'' \sqsubseteq w')$

[A3] $Int_\imath(\varphi) \rightarrow Bel_\imath(Int_\imath(\varphi))$

[SC3] $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(I(\imath)) \rightarrow \langle w,t,w''\rangle \in dom(B(\imath)) \wedge w' \in I_t^{w''}(\imath))$

[A4] $Des_\imath(\varphi) \rightarrow Bel_\imath(Des_\imath(\varphi))$

[SC4] $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in D(\imath) \rightarrow \langle w,t,w''\rangle \in B(\imath) \wedge w' \in D_t^{w''}(\imath))$

[A5] $Int_\imath(\varphi) \rightarrow Des_\imath(Int_\imath(\varphi))$

[SC5] $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(I(\imath)) \rightarrow \langle w,t,w''\rangle \in dom(D(\imath)) \wedge w' \in I_t^{w''}(\imath))$

A correspondência entre os axiomas da teoria formal e a abordagem semântica das lógicas BDI fuzzy é provada de forma análoga à prova da lógica BDI criada no capítulo anterior. Ou seja, os axiomas A1 e A2 são correspondentes às inter-relações SC1 e SC2 e isto é provado pelos corolários 7.3.1 e 7.3.2 (lógica BDI fuzzy versão I) e pelos corolários 7.3.4 e 7.3.5 (lógica BDI fuzzy versão II). Já as correspondências entre os axiomas A3 e SC3, A4 e SC4 e A5 e SC5 são garantidas pela definição 5.2.1 (versão I) e 7.3.12 (versão II). A prova destas correspondências, na versão I da lógica BDI fuzzy, é análoga à prova da proposição 5.3.1. As provas para a versão II são apresentadas a seguir.

Proposição 7.4.1 $\models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Bel_\imath(Int_\imath Opt(\varphi))$.

PROVA: Pela definição 7.3.12, $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(I(\imath)) \rightarrow \langle w,t,w''\rangle \in dom(B(\imath)) \wedge w' \in I_t^{w''}(\imath))$ e $M,V,w_t \models Int_\imath Opt(\varphi)$. A partir da semântica de Int, obtém-se que para todo $\langle w,t,x'\rangle \in dom(I(\imath))$ $M,V,x'_t \models Opt(\varphi)$. Pela condição semântica suposta haverá pelo menos um mundo w'' tal que $\langle w,t,w''\rangle \in dom(B(\imath))$ e que se relaciona com os mundos de intenções pertencentes ao $dom(I(\imath))$. Por consequência disto, tem-se que $M,V,w''_t \models Int_\imath(Opt(\varphi))$ e consequentemente $M,V,w_t \models Bel\imath(Int_\imath(Opt(\varphi)))$. Como, se $\forall w',w''.(w' \in I_t^w(\imath) \rightarrow w'' \in B_t^w(\imath) \wedge w' \in I_t^{w''}(\imath))$, então, para o mesmo $\langle M,V,w,t\rangle$, $M,V,w_t \models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Bel_\imath(Int_\imath Opt(\varphi))$. Logo, $\models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Bel_\imath(Int_\imath Opt(\varphi))$ ■

Corolário 7.4.1 *Seja φ uma fórmula de $L^{MF'}$. $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(I(\imath)) \rightarrow \langle w,t,w''\rangle \in dom(B(\imath)) \wedge w' \in I_t^{w''}(\imath))$ sse $\models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Bel_\imath(Int_\imath Opt(\varphi))$.*

Proposição 7.4.2 $\models^{k_{II}} Des_\imath(Opt(\varphi)) \rightarrow Bel_\imath(Int_\imath(Opt(\varphi)))$

PROVA: Prova análoga à prova da proposição 7.4.1. ■

Corolário 7.4.2 *Seja φ uma fórmula de $L^{MF'}$. $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in D(\imath) \rightarrow \langle w,t,w''\rangle \in B(\imath) \wedge w' \in D_t^{w''}(\imath))$ sse, $\models^{k_{II}} Des_\imath(Opt(\varphi)) \rightarrow Bel_\imath(Int_\imath(Opt(\varphi)))$.*

Proposição 7.4.3 $\models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Des_\imath(Int_\imath Opt(\varphi))$.

PROVA: Prova análoga à prova da proposição 7.4.1. ■

Corolário 7.4.3 *Seja φ uma fórmula de $L^{MF'}$. $\forall\langle w,t,w'\rangle, \langle w,t,w''\rangle.(\langle w,t,w'\rangle \in dom(I(\imath)) \rightarrow \langle w,t,w''\rangle \in dom(D(\imath)) \wedge w' \in I_t^{w''}(\imath))$ sse $\models^{k_{II}} Int_\imath Opt(\varphi) \rightarrow Des_\imath(Int_\imath Opt(\varphi))$.*

7.5 Considerações Finais

Mostra-se um bom número de possibilidades de construção de lógicas BDI fuzzy, todas, com diferentes semânticas (uma para cada caso a ser aplicado). Dessas, duas foram escolhidas para serem definidas. Foi, portanto, definido neste capítulo as abordagens semânticas de duas lógicas BDI fuzzy e uma teoria formal cuja corretude e completude são parcialmente provadas.

A partir das formalizações apresentadas aqui, tem-se agora aparatos suficientes para a formalização de modelos os quais funcionam como arquiteturas de sistemas multiagentes capazes de deliberar através de uma postura intencional (um exemplo de um modelo formal do raciocínio de um agente será visto no capítulo seguinte). Esta postura é uma extensão da teoria criada por Bratman, onde os agente tentam deliberar e satisfazer ações ou estados mentais a partir das suas crenças, desejos e intenções. Pois, tem-se aqui um sistema de inferência formal capaz de representar o raciocínio de um agente cognitivo de acordo com o que foi definido pela teoria de Bratman [Bra87] estendendo ainda este, no sentido de possibilitar o agente ter graus distintos de crenças, desejos e intenções.

Enfim, este capítulo dá oportunidade para vários trabalhos futuros iniciais, tais como: um estudo sobre as consequências da fuzzyficação de uma lógica BDI, analisando, assim, em estudos de casos como ela trataria um dado problema; diferenças entre as lógicas BDI fuzzy e análise de quais tipo de sistemas multiagentes elas seriam mais cabíveis; modelagem de sistemas de inferência modais fuzzy e de sistemas multiagentes, utilizando *AgentSpeak* ou JACK (estende a linguagem JAVA com um número de características BDI– Busetta, Howden, Ronnquist e Hodgson, em [PB00], utilizam esta linguagem para implementar um *cluster* de componentes de um agente BDI); definir noções e propriedades com relação à comunicação, cooperação, coordenação e argumentação de agentes que têm dúvidas e diferentes graus de desejos e intenções conflitantes ou até contraditórios ou não.

Capítulo 8

Estudo de Caso

Apresentar-se-á, neste capítulo, um estudo de caso, onde será aplicada a arquitetura interna de um agente BDI fuzzy. Esta arquitetura será fundamentada nas formalizações lógicas BDI e BDI fuzzy desenvolvidas neste trabalho. Podendo ser observado, também, superficialmente, como seria um sistema de inferência BDI fuzzy para tal aplicação.

Aplicação: O problema em questão a ser analisado sob as perspectivas BDI e BDI fuzzy, é o jogo "Captura" composto por três personagens: um monstro, um herói e uma princesa. A última está aprisionada em uma dada posição do tabuleiro. Ela deve ser alcançada pelo herói (ambos estarem na mesma posição). O herói tenta chegar a ela enquanto foge do monstro (personagem responsável pela captura da princesa), pois este deseja também capturar o herói. Cada personagem sabe a sua posição e a posição dos outros. E no tabuleiro, há posições vagas e posições bloqueadas, onde é possível apenas se posicionar sobre as vagas.

A cada instante de tempo, o monstro e o herói têm de se mover em uma das direções indicadas pelos pontos cardinais norte, sul, leste e oeste (não podem andar em diagonal). E é permitido apenas mover-se para posições adjacentes à atual.

A aplicação descrita nos parágrafos anteriores é apresentada na figura 8.1[1]:

Neste sistema multiagente aplicar-se-á às lógicas BDI e BDI fuzzy com o intuito de modelar formalmente o raciocínio do agente monstro, fazendo com que ele pretenda alcançar o herói e impeça este de chegar até a princesa.

[1]Os desenhos dos personagens foram retirados de [wen], [Mig] e [pri]

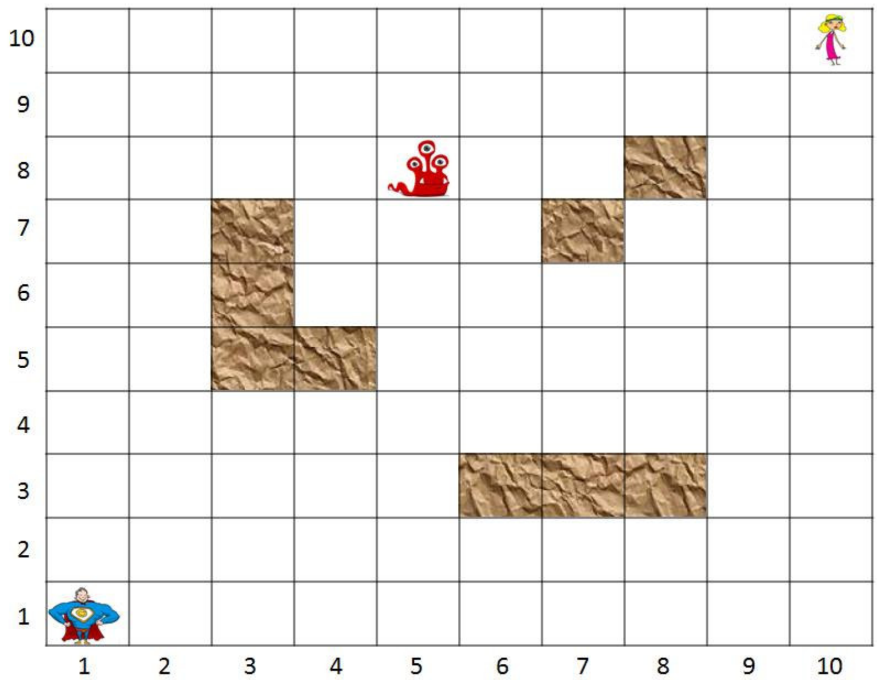


Figura 8.1: Configuração inicial do tabuleiro do jogo “Captura”.

8.1 Modelagem BDI

Nesta aplicação há os seguintes predicados:

- $Posicao(i, x, y)$, onde i é um agente, x é a posição de i no eixo x e y é a posição de i no eixo y. Daí, o predicado denota em qual posição (x,y) o agente i está,
- $Vago(x, y)$, denotando que a posição (x,y) é uma posição vaga,
- $Dist(x_{i_1}, y_{i_1}, x_{i_2}, y_{i_2})$, denotando a distância entre dois agentes a qual é calculada através da expressão $\sqrt{|x_{i_1} - x_{i_2}|^2 + |y_{i_1} - y_{i_2}|^2}$,
- $Direcao(i, pc, x, y)$, denotando a posição (x,y) – mais próxima, já que o personagem só pode andar uma posição por vez – de cada ponto cardinal (norte, leste, oeste e sul) em relação ao agente i, e
- $mover(i, pc)$, indica a direção do movimento de um dado agente.

Dessa forma, tem-se que os objetos ‘pc’, ‘x’ e ‘y’ pertencem ao domínio D_U enquanto $D_{Ag} = \{monstro, heroi, princesa\}$.

As intenções do monstro são:

1. $Inev(\Diamond(Posicao(heroi, x, y) \wedge Posicao(monstro, x, y)))$
2. $Inev(\Box(\neg(Posicao(heroi, x, y) \wedge Posicao(princesa, x, y))))$

O desejo do monstro é perseguir o herói até capturá-lo, o que vai mudar de natureza – ir em uma dada direção em um dado momento e movimentar-se para outro em outro dado momento – ao longo do tempo. Portanto, as crenças do monstro no início da partida são:

1. $Posicao(heroi, 1, 1) \wedge Posicao(monstro, 5, 8) \wedge Posicao(princesa, 10, 10)$
2. $\neg Vago(3, 7) \wedge \neg Vago(3, 6) \wedge \neg Vago(3, 5) \wedge \neg Vago(4, 5) \wedge \neg Vago(7, 7) \wedge \neg Vago(8, 8) \wedge \neg Vago(3, 7) \wedge \neg Vago(6, 3) \wedge \neg Vago(7, 3) \wedge \neg Vago(8, 3) \wedge Vago(1, 2) \wedge \ldots$
3. $Inev(Opt(Mover(norte)) \wedge Opt(Mover(sul)) \wedge Opt(Mover(leste)) \wedge Opt(Mover(oeste))$

O plano do agente monstro é então verificar se a posição dada pela direção de cada ponto cardinal está vaga e calcular a distância entre essa posição e a posição do herói. Daí, ele deve comparar tais distâncias e escolher a menor delas.[2]

Logo, segundo a lógica nova lógica BDI definida no capítulo 6 deste trabalho, no início do jogo, há a seguinte formalização:

$$
\begin{aligned}
M, V, w_t \models\ & Bel_{monstro}(Posicao(heroi, 1, 1) \wedge Posicao(monstro, 5, 8) \wedge Posicao(princesa, 10, 10)) \wedge \\
& Bel_{monstro}(\neg Vago(3, 7) \wedge \neg Vago(3, 6) \wedge \neg Vago(3, 5) \wedge \neg Vago(4, 5) \wedge \neg Vago(7, 7) \\
& \wedge \neg Vago(8, 8) \wedge \neg Vago(3, 7) \wedge \ldots) \wedge Bel_{monstro}(Inev(Opt(Mover(norte)) \\
& \wedge Opt(Mover(sul)) \wedge Opt(Mover(leste)) \wedge Opt(Mover(oeste)))) \\[1em]
M, V, w_t \models\ & Des_{monstro}(Inev(Opt(Mover(norte)) \wedge Opt(Mover(sul)) \wedge Opt(Mover(leste)) \\
& \wedge Opt(Mover(oeste))) \\[1em]
M, V, w_t \models\ & Des_{monstro}(Inev(\Diamond(Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y)) \\
& \wedge \Box(\neg(Posicao(heroi, x, y) \wedge Posicao(princesa, x, y))))) \\[1em]
M, V, w_t \models\ & Int_{monstro}(Inev(Opt(Mover(pc)))) \text{ onde } Direcao(pc, x, y) \text{ é uma posição a qual ele} \\
& \text{acredita ser vaga e ter a menor distância entre a mesma e a posição do herói.} \\[1em]
M, V, w_t \models\ & Int_{monstro}(Inev(\Diamond((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y))) \\
& \wedge \Box(\neg(Posicao(heroi, x, y) \wedge Posicao(princesa, x, y)))))
\end{aligned}
\tag{8.1}
$$

Na figura a seguir, é explicitado o modelo BDI e como ele progride através dos tempos e dos mundos com o intuito de uma melhor visualização do sistema BDI. As crenças, desejos e intenções apresentadas na figura são as mesmas explicitadas anteriormente:

[2] O plano do monstro, apresentado aqui, é um dos possíveis planos os quais o mesmo poderia adotar. Ele poderia, por exemplo aguardar o herói próximo à princesa e mover-se em direção a este caso o mesmo aloca-se próximo a ele.

- B1: Posições dos personagens
- B2: O que está vago
- B3: O monstro vai se mover
- D1: Mover-se
- D2: As posições do herói e do monstro eventualmente serão iguais e as do herói e da princesa sempre serão distintas
- I1: Mover em direção a um dos pontos cardinais
- I2: As posições do herói e do monstro eventualmente serão iguais e as do herói e da princesa sempre serão distintas (Idem 2)

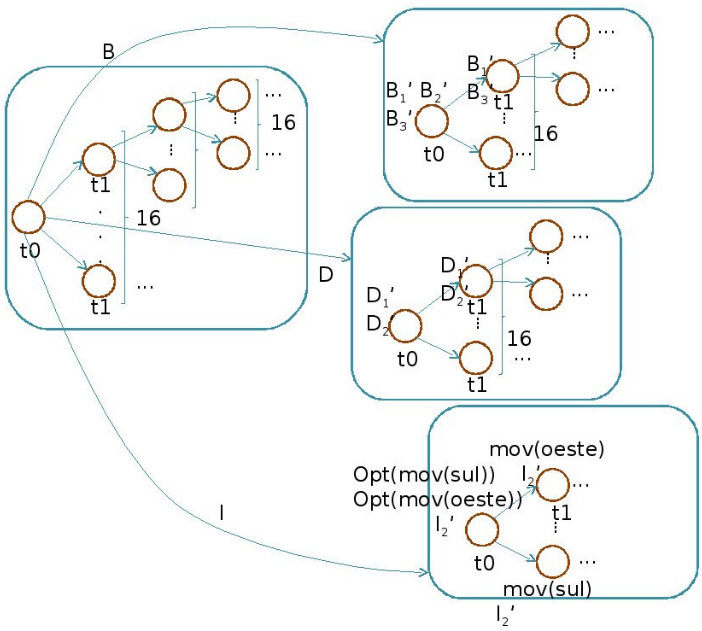


Figura 8.2: Modelo BDI do jogo "Captura".

Vê-se na figura que há 16 opções de situações. Tais opções são consequências das 4 possibilidades de movimento para cada agente (monstro e herói). Para cada situação, há um conjunto diferente de crença, desejo e intenção. B1, B3 e D1 serão iguais para todos os estados futuros e I2 e D2 terão sua manutenção segundo a estratégia de compromisso adotada pelo agente.

Por fim, a revisão das intenções deve ser feita a cada passo (após cada instante t). Isto fará com que o compromisso com a primeira intenção seja eliminado (supondo que ela seja satisfeita) e, dependendo do tipo de compromisso, as duas últimas intenções são mantidas até que: (1) o herói seja alcançado pelo monstro –

$Bel_{monstro}(((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y)))?)$ – (compromisso cego);

(2) até que 1 ou até que ele acredite que eventualmente irá satisfazê-la –

$Bel_{monstro}(\neg Opt(\Diamond((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y)) \vee (\neg\Diamond(Posicao(heroi, x, y)$ $\wedge Posicao(princesa, x, y))))))$ – (compromisso simples);

ou (3) até que 1 ou até que ele deseje que eventualmente irá satisfazê-la –

$Bel_{monstro}(\neg Opt(\Diamond((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y)) \wedge (\neg\Diamond(Posicao(heroi, x, y)$ $\wedge Posicao(princesa, x, y))))))$ – (compromisso mente-aberta).

8.2 Modelagem BDI Fuzzy

Caso haja o desejo de modelar as incertezas dos agentes quanto às suas crenças, desejos e intenções (com valores fuzzy), possibilita-se, então, aproximar ainda mais o sistema em questão de um sistema composto por agentes com características mais humanas. Tais como: visibilidade deficitária em função de bloqueios e da distância; e manipulação de desejos e intenções cujos graus de vontade de satisfazê-los não são – apenas – estritos, tornando possível formalizar a deliberação desses desejos e intenções.

As crenças do agente monstro envolvem sentenças relacionadas às posições vagas e bloqueadas e às posições ocupadas pelo herói. Os graus de verdade dessas sentenças podem ser definidos conforme a distância entre a posição do agente e a posição do que ele crê estar vago (ou bloqueado) e entre a posição dele (monstro) e o agente herói; de forma que quanto maior a distância entre ele e a posição do objeto de crença, menor a sua crença nesta informação. E se houver uma posição bloqueada entre as posições (do monstro e o objeto do predicado da sentença), a crença do monstro seria reduzida (em 0,4 grau, por exemplo).

Assim, é possível construir uma situação em que o agente tem o grau de certeza de sua crença reduzida em função da sua visão, seja por causa da distância ou porque a mesma está bloqueada. O grau de crença da posição da princesa é considerado '1', pois foi o próprio monstro que a aprisionou.[3]

Portanto, no início do jogo, as crenças do monstro são formalizadas da mesma forma a qual foi definida para a modelagem BDI:

$$\begin{aligned}
&Bel_{monstro}(Posicao(heroi, 1, 1) \wedge Posicao(monstro, 5, 8) \wedge Posicao(princesa,\\
&10, 10)) \wedge Bel_{monstro}(\neg Vago(3, 7) \wedge \neg Vago(3, 6) \wedge \neg Vago(3, 5) \wedge \neg Vago(4, 5)\\
&\wedge \neg Vago(7, 7) \wedge \neg Vago(8, 8) \wedge \neg Vago(3, 7) \wedge \ldots) \wedge (Inev(Opt(Mover(norte))\\
&\wedge Opt(Mover(sul)) \wedge Opt(Mover(leste)) \wedge Opt(Mover(oeste))))
\end{aligned} \tag{8.2}$$

Observe que os graus de verdade dessa crença podem variar com respeito à lógica BDI fuzzy utilizada. Por exemplo, caso fosse utilizada a versão I, a crença em tal sentença dependeria unicamente do valor de cada sub-fómula no mundo de crença correspondente. Portanto, se $V_t^{w'}(Posicao(heroi, 1, 1)) = 0,58$ onde

[3] A crença do monstro com relação a posição da princesa também poderia ter seu grau de certeza reduzido em função do tempo, representando assim a falta de memória do monstro. E caso fosse modelada as crenças do herói poder-se-ia definir algum super poder do mesmo, como por exemplo: "visão de raio-x".

$w' \in B_t^w(monstro)$ então $V(Bel_{monstro}(Posicao(heroi, 1, 1)) = 0,58$ e o grau da crença na fbf descrita pela equação 8.2 será dado por uma t-norma composta pelos valores de cada sub-fómula.

Os desejos do monstro diminuirão de acordo com a diferença entre as distâncias referente às posições do monstro e do herói e às posições do herói e da princesa. Ou seja: quão mais próximo o herói esteja do monstro, maior o desejo sobre esta sentença; e quão mais longe estiver o herói da princesa, maior o desejo sobre esta sentença. Já os desejos com relação ao movimento do monstro, esses são valorados de forma que o movimento a torná-lo mais próximo do herói receba a maior graduação ('1'), enquanto os outros valores serão decrescidos à medida que a direção não o aproxime tanto do herói, até o ponto em que mais o distancie. Assim, os graus dos desejos do monstro no estado inicial do jogo terão os seguintes valores:

$$\begin{aligned}
&V_t^w(Opt(Mover(norte))) = 0\\
&V_t^w(Opt(Mover(sul))) = 1\\
&V_t^w(Opt(Mover(leste))) = 0,333\ldots\\
&V_t^w(Opt(Mover(oeste))) = 1\\
&V_t^w(Inev(\Diamond((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y))))) = 0,9\\
&V_t^w(Des_{monstro}(Inev(\Diamond(\neg(Posicao(heroi, x, y) \wedge Posicao(princesa, x, y)))) = 1
\end{aligned} \tag{8.3}$$

Note que o último desejo recebeu um grau de certeza máximo, enquanto o penúltimo desejo foi modelado com grau próximo de '1'. Tal valoração foi assim definida devido à preferência do monstro em não permitir que o herói salve a princesa, a capturar o herói. Uma vez que se o primeiro caso ocorrer, o monstro terá ninguém sob o seu domínio. Note também que ao passar do tempo, esses desejos podem diminuir em função da falta de expectativa de conseguir alcançá-los, influenciando então diretamente no compromisso com ambos. Tal diminuição pode ser modelada através de regras de inferência, considerando a distância entre os agentes – quão mais distante o monstro estiver do herói e da princesa e mais próximo o herói da princesa, mais difícil se tornará a satisfação dos dois últimos desejos explicitados em 8.3. Por fim, observe que a estipulação de valores de verdade deve ser feito pelo sistema. O mesmo também é responsável por determinar quais valores são considerados verdadeiros e quais valores são considerados falsos. No caso de um sistema de inferência BDI fuzzy a atribuição dos graus de pertinência dos desejos seriam definida pelo especialista no problema (responsável por modelar o sistema).

Em consequência da definição das relações de acessibilidade (seguindo a abordagem semântica) e do axioma A2 da teoria formal BDI/BDI fuzzy apresentada neste trabalho, o grau de verdade das intenções do monstro são:

$$\begin{aligned}
&V_t^w(Int_{monstro}(Inev(Opt((Mover(sul) \wedge Opt(Mover(oeste)))))) = 1\\
&V_t^w(Int_{monstro}(Inev(\Diamond((Posicao(heroi, x, y) \wedge Posicao(Monstro, x, y)))))) = 0,9\\
&V_t^w(Int_{monstro}(Inev(\Box(\neg(Posicao(heroi, x, y) \wedge Posicao(princesa, x, y)))))) = 1
\end{aligned} \tag{8.4}$$

Observe que houve a consideração das opções $(Mover(sul)$ e $Mover(oeste)$ como uma das intenções do monstro. O monstro no "mundo real" pode tomar qualquer uma das opções como caminhos a serem executados, ou seja, ele pode deliberar qualquer uma das ações sem prejuízo de desempenho. Neste momento, fica visível a possibilidade de se construir um sistema de inferência BDI fuzzy que seja capaz de determinar qual a ação a ser realizada. E desta forma, capacitar um sistema BDI a formalizar a deliberação e a reconsideração de intenções de um agente BDI, o que não é possível para nenhuma das formalizações lógicas de arquitetura interna de agentes cognitivos vistas aqui neste trabalho.

As escolhas de um conjunto de intenções, dentre as opções de desejos do agente, é conseguida através da escolha dos desejos com maiores graus de certeza. Logo, um sistema de inferência BDI fuzzy pode ser construído para este estudo de caso. Tal sistema teria como entrada conjuntos fuzzy de crenças e desejos onde os graus de pertinência de cada desejo e crença seriam dados a partir da distância entre o monstro e o objeto da atitude mental (posição ou agente herói) – segundo as regras já explicitadas em parágrafos anteriores. Daí, a partir dos axiomas da teoria formal BDI, definida neste trabalho, e algumas regras de inferência (Modus Ponens), obter-se-ia um conjunto de intenções fuzzy de saída.

Observação 8.2.1 *A valoração fuzzy das fórmulas envolvendo crenças, desejos e intenções do agente monstro não são especificadas em função do grau de certeza da sub-fórmula dessas atitudes. Isto foi feito para que a verificação da formalização* BDI *fuzzy fosse feita independente da lógica* BDI *a ser utilizada, tornando hábil a aplicação de qualquer formalização* BDI *fuzzy no estudo do caso.*

8.3 Considerações Finais

Neste capítulo foi analisado o raciocínio de um agente cognitivo cuja formalização do mesmo foi fundamentada na lógica BDI e nas lógicas BDI fuzzy (independente de qual seja a lógica BDI fuzzy). Com este exemplo de sistema multiagente, apesar de bastante simples, pôde ser observado que com a utilização de uma fundamentação BDI fuzzy há um grande enriquecimento do sistema, com relação ao número de possibilidades a serem representadas. Além da possibilidade da determinação de qual intenção tomar em função das opções de desejos do agente em um dado instante. Evidencia-se, portanto, que um conjunto de intenções fuzzy de saída pode ser gerado em função das entradas: conjuntos fuzzy de crenças e conjuntos fuzzy de desejos. Onde as regras de inferência contidas neste sistema de inferência BDI fuzzy levariam em consideração os próprios axiomas (inter-relações entre as atitudes) da teoria formal.

Entretanto, tais vantagens têm um preço. Percebe-se que ao estender um sistema multiagente fundamentado na lógica BDI para um sistema fundamentado na lógica BDI fuzzy, a complexidade de tal sistema aumenta em decorrência da maior quantidade e complexidade da atribuição de valores verdades – valoração das relações entre mundos e estados de tempo caso fosse usada a versão II – e da maior quantidade de situações que o agente deve analisar para tomar sua decisão – não há mais deliberações aleatórias de intenções conflitantes; a não ser que elas não prejudiquem o desempenho do agente.

Há inúmeras possibilidades de aumentar a complexidade do problema, aumentando o número de ações, o número de agentes – podendo então modelar, por exemplo, a cooperação entre eles, a coordenação de suas ações e até mesmo um sistema de valores de troca (vide [RCB03] para mais detalhes) – ou possibilitando um agente realizar parcialmete uma ou mais ações (podendo ficar por exemplo entre uma casa e outra do tabuleiro). Enfim, há uma grande quantidade de situações a serem criadas as quais podem (todas) ser representadas por um sistema formal BDI fuzzy.

Por fim, pôde ser observado a relação de $\cap_{sub}$ entre os desejos e as crenças. E a relação de $\subseteq_{sup}$ dos desejos com respeito às intenções.

Capítulo 9

Conclusão

Nesta dissertação, foram definidas lógicas BDI fuzzy cujas principais aplicações são as formalizações de arquiteturas internas de agente cognitivos, em especial, agentes BDI. E como tais lógicas são usadas como fundamentações de sistemas multiagentes, então apresentou-se aqui uma introdução sobre sistemas multiagentes, bem como, os assuntos iniciais sobre agentes e agentes cognitivos. Viu-se também as principais formalizações lógicas de arquiteturas de agentes BDI existentes e explicou-se o porquê da utilização da teoria BDI argumentando sua fundamentação filosófica e suas várias implementações computacionais.

Há na literatura duas formalizações BDI (fundamentadas unicamente na teoria de Bratman), uma utiliza a abordagem semântica ([Woo00a]) e outra utiliza a abordagem formal (teoria formal) – [RG91, RG93]. Entretanto, preferiu-se fazer algumas modificações focando principalmente na teoria de Bratman. E com isso ter um sistema lógico formal mais apto a representar a cognição humana com respeito à deliberação de ações (formalizando assim, a tese de assimetria e adequando-a aos conceitos de, e relações entre crenças, desejos e intenções sugeridos por Bratman).

Esses trabalhos sofreram uma rigorosa análise para daí se tentar retirar algumas ambiguidades, presentes em tais textos as quais foram apontadas no quinto capítulo deste trabalho. A solução escolhida para tratar tais ambiguidades foi a mais simples, pois a notação da lógica BDI já é bastante "pesada" e complicada para transformá-la em uma lógica BDI fuzzy de alta ordem. No entanto, seria interessante implementar esta lógica e fazer comparações a fim de analisar a aplicabilidade de ambas.

Logo, tem-se que a lógica BDI definida neste trabalho é a lógica BDI– com algumas provas das correpondências entre os axiomas e a abordagem semântica da mesma – mais próxima da teoria de Bratman que se tem atualmente. E há a definição de várias lógicas BDI fuzzy – extensões dessa lógica BDI– as quais, consequentemente, (conforme provado), também têm a tese de assimetria de Bratman inerente à sua definição.

Como trabalho futuro direto desta dissertação, há a prova dos teoremas de completude e corretude das lógicas BDI e BDI fuzzy, o que seria um trabalho árduo e de alta complexidade. E a implementação de um sistema de inferência BDI fuzzy e aplicá-lo no estudo de caso apresentado, com a finalidade de formalizar a deliberação e reconsideração de intenções. Tais extensões do presente trabalho são importantes, pois dariam a certeza de que os problemas apontados nas lógicas anteriores foram solucionados, além de se estar lidando com uma lógica completa e correta; e porque nenhuma formalização lógica BDI, até então, conseguiu implementar tais funções. Nesta dissertação, mostrou-se como este sistema de inferência fuzzy seria feito, mas não houve a definição nem a aplicação do mesmo em um sistema multiagente.

Referências Bibliográficas

[All90] J. Allen. Two views of intention: Comments on bratman and on cohen and levesque. In M.E. Pollack P.R. Cohen, J. Morgan, editor, *Intentions in Communication*. MIT Press, Cambridge, MA, 1990.

[ATV80] C. Alsina, E. Trillas, and L. Valverde. On non-distributive logical connectives for fuzzy set theory. *Busefal*, 3:18–29, 1980.

[BA07] B. C. Bedregal and B. M. Acióly. *Introdução à Lógica Classica para Ciência da Computação*. 2007.

[Bac04] M. Baczynski. Residual implications revisited. notes on the smets-magrez. *Fuzzy Sets and Systems*, 145(2):267–277, 2004.

[BBS03] H. Bustince, P. Burillo, and F. Soria. Automorphism, negations and implication operators. *Fuzzy Sets and Systems*, 134:209–229, 2003.

[BC08] B. C. Bedregal and A. P. Cruz. A charachterization of classic-like fuzzy semantics logical as a propositional fuzzy logic. *Logical Journal of IGPL*, 16:357–370, 2008.

[BHW07] R. H. Bordini, J. F. Hübner, and M. Wooldridge. *Programming Multi-agent Systems in AgentSpeak using Jason*. WILEY, England, 2007.

[Bin92] K. Binmore. *Fun and Games: A Text on Game Theory*. D. C. Hath and Company, Lexington, MA, 1992.

[BIP88] M. Bratman, D. J. Israel, and M. E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[BM02] R.H. Bordini and A.F. Moreira. Proving the assymetry thesis principles for a BDI agent-oriented programming language. *Electronic Notes in Theoretical Computer Sciences*, 70(5):108–125, 2002.

[Bra84] M. E. Bratman. Two faces of intention. *Philos. Rev.*, 93:375–405, 1984.

[Bra87] M. E. Bratman. *Intentions, Plans and Pratical Reason*. The David Hume Series. CSLI, Cambridge, Mass, 1987.

[CB07] A. P. Cruz and B. C. Bedregal. Quando a teoria fuzzy é paraconsistente? In *Congress of Logic Applied to Technology*, Santos, Outubro 2007.

[CL87] P. R. Cohen and H. J. Levesque. Persistence, intention and commitment. In *Reasoning about Action and Plans*, pages 297–340, Los Altos, Cal, 1987. Morgan Kaufman.

[CL90] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[Cre70] M. J. Cresswell. Classical intensional logic. *Theoria*, 36:347–372, 1970.

[Cre72] M. J. Cresswell. Intensional logics and logical truth. *J. Philosophical Logic*, 1:2–15, 1972.

[Cre73] M. J. Cresswell. *Logic and Languages*. Methuen, London, 1973.

[Den87] D. C. Dennett. *The Intentional Stance*. The MIT Press, Cambridge, MA, 1987.

[Den97] D. C. Dennett. *Tipos de Mentes: rumo a uma compreensão da consciência*. Rocco, Rio de Janeiro, 1997.

[dKLW97] M. d'Inverno, D. Kinny, M. Luck, and M. Wooldridge. *Intelligent Agents IV (LNAI Volume 1365)*, pages 155–176. Springer-Verlag, Berlin, Germany, 1997.

[ES89] E.A. Emerson and J. Srinivasan. Branching time temporal logic. In *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency, School/Workshop*, pages 123–172, London, UK, 1989. Springer-Verlag.

[Fer86] A. B. H. Ferreira. *Novo Dicionário Aurélio da Língua Portuguesa*. Editora Nova Fronteira, 2ª edition, 1986.

[Fit91] M. Fitting. Many-valued modal logics. *Fundamenta Informaticae*, 15:235–254, 1991.

[Fod91] J. C. Fodor. On fuzzy implication operators. *Fuzzy Sets and Systems*, 42:293–200, 1991.

[FR94] J. C. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publisher, 1994.

[Fre79] G. Frege. *Begriffsschrift*. 1879. English translation in HEIJENHOORT (1967).

[Gal88a] J. R. Galliers. A strategic framework for multiagent cooperative dialog. In *Proocedings of Eighth European Conference on Artificial Intelligence (ECAI-88)*, pages 415–420, Munich, Federal Republic of Germany, 1988.

[Gal88b] J. R. Galliers. *A theoretical framework for multiagent cooperative dialog, Acknowledging Multi-Agent Conflict*. PhD thesis, Open University, UK, 1988.

[Gar00] J. Garson. *Stanford Encyclopedia of Philosophy*. Stanford University, CA, February 2000. Avaliable in http://plato.stanford.edu/.

[Gol92] R. Goldblat. *Logics of Time and Computation*. Center for the Study of Language and Inf, 2 edition, September 1992.

[GPP+] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge. The belief-desire-intention model of agency. Disponível em www.cs.huji.ac.i/∼imas/readings/atal98b.pdf.

[Gär88] P. Gärdenfors. *Knowledge in Flux*. The MIT Press, Cambridge, MA, 1988.

[GR02] L. Godo and R. Rodríguez. Graded similarity based semantics for nonmonotonic inference. *Ann. Math Artificial Intelligence*, 24:89–105, 2002.

[Haa02] S. Haack. *Filosofia das Lógicas*. Editora Unesp, São Paulo, 2002. Translated by Cezar Montari and Luiz Dutra.

[HH96] P. Hájek and D. Harmancová. A many-valued modal logics. In *Proceedings of IPMU'96*, pages 1021–1024, 1996.

[Hin75] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophy*, 4:475–484, 1975.

[Háj98] P. Hájek. *Metamathematics of fuzzy logic(Trends in Logic)*, volume 4. Kluwer Publishers, Dordrecht, 1998.

[HN02] R. Horcik and M. Navara. Validation sets in fuzzy logics. *Kybernetika*, 38(3):319–326, 2002.

[HvdHMW02] P. Harrebstein, W. van der Hoek, J-J Meyer, and C. Wittenven. On modal logic interpretations for games. In *Proocedings of Fifteenth European Conference on Artificial Intelligence (ECAI-2002)*, pages 28–32, Lyon, France, 2002.

[HW03] W. Hoek and M. Wooldridge. Towards a logic of rational agency. In *L. J. of the IGPL*, volume 11, pages 135–159. Oxford University Press, 2003.

[Jer92] N. R. Jernnings. Towards a cooperation knowledge level for collaborative problem solving. In *Proocedings of Tenth European Conference on Artificial Intelligence (ECAI-92)*, pages 224–228, Vienna, Austria, 1992.

[KG91] D. Kinny and M. Geogerff. Commitment and effectiveness of situated agents. In *Proceddings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 82–88, Sidney, Australia, 1991.

[KK62] W. Kneale and M. Kneale. *O Desenvolvimento da Lógica*. The Clarendon Press, Oxford, 3th edition, 1962. Translated by M. S. Lourenço.

[KMP00] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Publisher, Dordrecht, 2000.

[KN99] E. P. Klement and M. Navara. A survey on different triangular norm-based fuzzy logics. *Fuzzy Sets and Systems*, 101:241–251, 1999.

[LCN90] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes. On acting together. In *Proocedings of Eighth National Conference on Artificial Intelligence (AAAI-90)*, pages 94–99, Boston, MA, 1990.

[Les03] J. M. Leski. Insensitive learning techniques for approximate reasoning system. *International Journal of Computational Cognition*, 1(1):21–77, 2003.

[Mig] Miguel. Disponível em http://miguelblogportugal.blogspot.com/2008/04/de-monstro-monstrinho.htm. Acessado em 19.08.08.

[Moo77] R. C. Moore. Reasoning about knowledge and action. In *Proceedings of the Fifth International Joint Conference on Artificial INteligence (IJCAI-77)*, Cambridge, MA, 1977.

[Mor00] B-A Mordechai. *Mathematical for Computer Science*. Springer, London, 2000.

[MW99] A. Rao M. Wooldridge, editor. *Foundations of Rational Agency*. Kluwer Academic Publishers, Boston, MA, 1999.

[Nav99] M. Navara. Characterization of measures based on strict triangular norms. *J. Math. Anal. Appl.*, 236:370–383, 1999.

[NM44] J. Von Newman and O. Morgenstern. *Theory of games and Economic behaviors*. Princeton University Press, Princeton, NJ, 1944.

[Non01] G. Nonanno. *Branching time logic, perfect information games and backward induction*, volume 36, chapter Games and Economic Behavior, pages 57–73. Havard University Press, July 2001.

[OR94] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, Cambridge, MA, 994.

[Par85] Parikh. The logic of games and its applications. In M. Karpinksy and J. van Leewen, editors, *Topics in the Theory of Computation*, Amsterdam, The Netherlands, 1985. Elsevier Science Publishers B.V.

[Pau02] M. Pauly. A modal logic for coalition power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[PB00] R. Ronnquist A. Hodgson P. Busetta, N. Howden. Structuring bdi agents in functional clusters. In N.R. Jennings and Y. Lespérance, editors, *Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, pages 277–289, Berlin, 2000. Springer-Verlag.

[PL87] M. P.Georgeff and A. L. Lansky. Reactiving reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 677–682, Seattle, WA, 1987.

[pri] Disponível em www.drawingcoach.com/cartoon-princess.html. Acessado em 19.08.08.

[Rao96a] A. S. Rao. Agentspeak(L): BDI agentes speak out in a logical computable language. In W. Van de Velde and J. W. Perram, editors, *Agents Breakin Away: Proceedings of the Seventh European Workshop on Modelling Autonomous Agents in a Multiagent World (LNAI Volume 1038)*, pages 42–55, Berlin, Germany, 1996. Springer-Verlag.

[Rao96b] A. S. Rao. *Intelligent Agents II: Agent Theories, Architectures and Language (IJCAI Volume 1037)*. Springer-Verlag, 1996.

[RCB03] M. R. Rodrigues, A. C. R. Costa, and R. H. Bordini. A system of exchange values to support social interactions in artificial societies. In *Autonomous Agents and Multiagent Systems - AAMAS 2003*, Melbourne, Julho 2003. Disponível em http://rocha.ucpel.tche.br/valores/.

[RG91] A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In Fikes R. Sandewall E. Allen, J., editor, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, KR91*, San Mateo, CA, february 1991. Morgan Kaufmann.

[RG93] A. S. Rao and M. P. Georgeff. Intentions and rational commitment. In *Proceedings od the First Pacific Rim International Conference on Artificial Intelligence*, Nagoya, Japan, October 1993.

[RG98] A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3):293–343, 1998.

[RGG96] R. Rodríguez, P. Garcia, and L. Godo. Using fuzzy similarity relations to revise and update a knowledge base. *Mathware Soft Comput.*, 3:357–370, 1996.

[RK93] D. Ruan and E.E. Kerre. Fuzzy implication operators and generalized fuzzy methods of cases. *Fuzzy Sets and Systems*, 54:23–37, 1993.

[Roc07] A. M. Rocha. Estudo da bi-implicação nebulosa. Trabalho de graduação disponível em http://www.dimap.ufrn.br/∼bedregal/students.html, Junho 2007.

[Rus03] B. Russel. *The Principles of Mathematics*. 1903.

[Sea83] J.R. Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, New York, 1983.

[SS63] B. Schweizer and A. Sklar. Associative functions and abstract semigroups. *Publicationes Mathmaticae Debrecen*, 10:69–81, 1963.

[Sti83] S. P. Stich. *From Folk Psycology to Cognitive Science*. The MIT Press, Cambridge, MA, 1983.

[Var86] M. Y. Vardi. On epistemic logic and logical omniscience. In J. Y. Halpern, editor, *Proceedings of the first Conference on Theorethical Aspects of Reasoning about Knowledge*, pages 293–306, San Mateo, California, 1986. Morgan Kaufmann Publishers.

[vLvdHM98] B. van Linder, W. van der Hoek, and J-J. Meyer. Formalizing abilities and opportunities of agents. *Fundameta Informaticae*, 34(1,2):53–101, 1998.

[wen] weno. Disponível em www.weno.com.br/blog/super_heroi.gif. Acessado em 19.08.08.

[WJ95] M. Wooldridge and N. R. Jernnings. Intelligent agents: Theory and practice. *The Knowledge Engennering Review*, 10(2):115–152, 1995.

[Woo00a] M. Wooldridge. *Reasoning About Rational Agents*. The MIT Press, Cambridge, MA, 2000.

[Woo00b] M. Wooldridge. *Reasoning About Rational Agents*, chapter 7, pages 127–134. The MIT Press, Cambridge, MA, 2000.

[Woo02a] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, LTD, London, UK, 2002.

[Woo02b] M. Wooldridge. *An Introduction to MultiAgents Systems*, chapter 8, pages 164–168. John Wiley & Sons, LTD, London, UK, 2002.

[Yag83] R. R. Yager. On the implication operator in fuzzy logic. *Information Science*, 31:141–164, 1983.

[Yag04] R. R. Yager. On some new classes of implication operators and their role in approximate reasoning. *Information Science*, 167:193–216, 2004.

[Zad65] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8, 1965.

[ZSC04] Z. Zhang, Y. Sui, and C. Cao. Fuzzy reasoning based on propositional modal logic. In *Proceedings of the Fourth Internat. Conf. on Rough Sets and Current Trends in Computing (RSCTC2004)*, volume 3066, pages 109–115, Berlin, 2004. Springer.

[ZSCG06] Z. Zhang, Y. Sui, C. Cao, and W. Guohua. A formal fuzzy reasoning system and reasoning mechanism based on propositional modal logic. *Theoretical Computer Science*, 368:149–160, 2006.