UM ESTUDO DO CONHECIMENTO: ALGUMAS ABORDAGENS PARA A SUA FORMALIZAÇÃO

Eliana Silva de Almeida

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

Sheila Regina Murgel Veloso

Profa. Sheila Regina Murgel Veloso, D.Sc.
(Presidente)

Sueli Mendes

Profa. Sueli Bandeira Teixeira Mendes, Ph.D.

Alfredo Vidal de Carvalho

Prof. Luis Alfredo Vidal de Carvalho, D.Sc.

Edward Hermann Haeusler

Prof. Edward Hermann Haeusler, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

NOVEMBRO DE 1991

ALMEIDA, ELIANA SILVA DE

Um Estudo do Conhecimento: Algumas Abordagens para a sua Formalização [Rio de Janeiro] 1991 viii, 182 p., 29,7 cm (COPPE/UFRJ, M. Sc., Engenharia de Sistemas e Computação, 1991)

Tese-Universidade Federal do Rio de Janeiro, COPPE

1. Formalização do Conhecimento 2. Raciocínio não-monotônico 3. Revisão de crença 4. Especificação de Sistemas Distribuídos I.COPPE/UFRJ II.TÍTULO (série)

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências (M. Sc.).

UM ESTUDO DO CONHECIMENTO: ALGUMAS ABORDAGENS PARA A SUA FORMALIZAÇÃO

Eliana Silva de Almeida

Novembro de 1991

Orientadora: Profa. Sheila Regina Murgel Veloso

Programa: Engenharia de Sistemas e Computação

São abordadas diversas formalizações lógicas para as diferentes noções de conhecimento, visando as aplicações no campo da Inteligência Artificial em geral.

Serão analisados o modelo dos "mundos possíveis", a linguagem que formaliza este modelo e uma ferramenta que dá a semântica da linguagem chamada "estrutura Kripke".

Algumas abordagens que formalizam o conhecimento serão estudadas, enfocando-se o problema da "omnisciência lógica" e o raciocínio não-monotônico.

Como aplicação é apresentada uma formalização de um Sistema Distribuído, onde dois problemas, "o ataque coordenado" e o "problema dos maridos infiéis", são analisados utilizando uma abordagem baseada na lógica do conhecimento.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Master of Science (M. Sc.).

A STUDY OF KNOWLEDGE: SOME APPROACHES TO FORMALIZATION

Eliana Silva de Almeida

November, 1991

Thesis Supervisor: Profa. Sheila Regina Murgel Veloso

Department: Systems Engineering and computing

Some logical formalizations to the differents notions of knowledge are approched, seeking applications in the field of Artificial Intelligence in general.

It will be analyzed the model of "possible words", the language that formalizes this model and a framework that gives semantic to language called "Kripke structure".

Some approaches that formalizes knowledge will be studied, emphasizing the problem of "omniscience logic" and nonmonotonic reasoning.

As an application is presented a formalization of the distributed system, where two characteristics problems, "the coordinated attaches" and "the problem of cheating husbands", are analyzed using an approach based on the logic of knowledge.

Para Dismaldo

e Élida

AGRADECIMENTOS:

- A minha orientadora, Profa. Sheila Regina Murgel Veloso, pela dedicação e pela segurança com que orientou esta tese.

- Aos meus pais, Disnaldo e Élida, aos meus irmãos Gustavo e Paula e a minha querida "vó" Percília, pela preocupação e por todo amor e carinho que sempre me transmitiram.

- Aos colegas do Depto. de Computação da Universidade Federal Fluminense por terem me incentivado a concluir este trabalho.

- Aos amigos Edson, Célia e Fernando pelos agradáveis "estudos em grupo" que participamos.

- Ao amigo João Carlos por ter feito a revisão deste trabalho e pelas dúvidas esclarecidas e ao amigo Alexandre pelos artigos conseguidos.

- Ao casal Zadir e Therezinha e aos queridos amigos Henrique, Patrícia e Dill, pelos "almoços em família" e pelos momentos agradáveis que passamos juntos.

- As amigas Ada Cristina, Andressa, Anna Karina, Emília, Joana, Mônica, Tatiana e Zenaide por todo o carinho e atenção.

CAPÍTULO I

INTRODUÇÃO

A proposta deste trabalho é estudar o conhecimento utilizando formalizações lógicas. A motivação para este estudo se encontra na possibilidade de, através de fatos que são conhecidos por um agente, atingir padrões de "raciocínio" que se assemelham ao raciocínio humano.

Apesar do estudo do conhecimento ser bem antigo entre a comunidade filosófica (HINTIKKA, 1962), a idéia de sua formalização utilizando uma linguagem lógica é bem mais recente. Este estudo só foi possível quando foi considerada a hipótese de se ter um sistema constituido por um conjunto de fatos "conhecidos" em um certo "mundo", os quais seriam manifestados de acordo com o comportamento do sistema. Várias aplicações são encontradas em áreas como Banco de Dados, Robótica, Inteligência Artificial e, em particular, na especificação de Sistemas Distribuídos.

Dado um agente A, chama-se de Base de Conhecimento o conjunto de fatos que são conhecidos por A. Será então estabelecido o que irá significar "um agente conhece um fato α": dado uma base de conhecimento Γ, para o agente "descobrir" se conhece α, intuitivamente, ele irá "raciocinar" com Γ, fazendo uso de um procedimento de inferência, de modo que, caso α possa ser deduzido de Γ, então ele será um fato conhecido do agente. É possível estender a noção de conhecimento de um agente para um grupo de agentes, onde novas noções como "conhecimento implícito", "conhecimento comum", etc, podem ser

introduzidas.

Com este intuito, uma linguagem lógica é utilizada, devendo ser poderosa o suficiente para expressar as possíveis noções de conhecimento. Além da linguagem, a lógica do sistema deverá ser tal que todas as inferências sejam intuitivamente corretas. A semântica formal é então escolhida de modo que o procedimento de inferência seja "correto" e "completo" em relação a mesma. Para isto, são consideradas as propriedades que o conhecimento deve ter, que identificarão o que é e o que não é conhecido. Estas propriedades permitirão identificar não apenas noções de conhecimento mas também as noções de crença.

O modelo clássico para o conhecimento ou crença é o modelo dos "mundos possíveis", cuja idéia intuitiva nos diz que, apesar do agente possuir um conjunto de fatos que são verdadeiros em um "mundo", o seu "mundo real", existirão outros "mundos", os mundos possíveis, em que este agente poderia possuir este mesmo conjunto de fatos. O agente diz então que conhece um fato α se α é verdadeiro em todos os mundos possíveis para ele.

Segundo a literatura, um problema com a semântica dos mundos possíveis diz que ela não é muito apropriada para modelar o raciocínio humano. Na realidade, este modelo sofre do problema chamado de "omnisciência lógica". Dizer que um agente é logicamente omnisciente significa dizer que ele conhece todos os fatos válidos e o seu conhecimento é fechado sobre a implicação, isto é , se o agente conhece um fato α e também tem o conhecimento de $(\alpha \supset \beta)$, então o agente deve também conhecer β. Na vida real não são encontradas pessoas com as características deste agente, ou

seja, as pessoas não são completamente omniscientes. Assim, vários formalismos lógicos foram propostos para tratar a falta de ominisciência lógica. Cada um deles atribuia um certo motivo para a existência do problema e a solução era então atingida com bases neste motivo.

Um outro problema a considerar está relacionado ao conjunto de fatos que são conhecidos pelo agente pois, é a partir deste conjunto que o agente pode inferir novos fatos (dedução). Ocorre que, nos formalismos mais comuns, baseados na lógica clássica, quando novos fatos são inferidos, estes são preservados mesmo que outros fatos sejam adicionados ao conjunto, ou seja, a lógica clássica é monotônica. Mais uma vez, este tipo de raciocínio não corresponde ao raciocínio humano, que possui características não-monotônicas: as pessoas, durante o processo de dedução, baseiam-se em um conjunto de fatos a respeito do mundo, o qual pode conter informações incompletas que, quando alteradas, poderão mudar suas decisões ou conclusões.

Dentre as formalizações que modelam este raciocínio não-monotônico, o nosso interesse está voltado para aquelas baseadas em uma lógica que capture a noção de conhecimento ou crença. Nestas formalizações, ao considerar a base de conhecimento do agente Γ, esta irá representar "tudo o que é conhecido", mas, quando certas hipóteses sobre a capacidade introspectiva do agente são consideradas, novos "conhecimentos" podem aparecer. Como exemplo, suponha Γ contendo uma fórmula α. As fórmulas conhecidas pelo agente serão todas aquelas deduzidas de Γ mas, se β for uma outra fórmula, por introspecção, o agente não irá conhecer β e

terá o conhecimento de que ele não conhece β, muito embora este fato não seja deduzido de Γ. Caso β venha pertencer a Γ, então β será uma fórmula conhecida pelo agente e o conjunto de fórmulas conhecidas poderá não só ser alterado como novas fórmulas poderão ser acrescentadas e não mais teremos que o agente não conhece β.

Neste trabalho o que se pretende é mostrar como a noção de conhecimento e crença pode ser formalizada. São apresentadas diversas abordagens, que são utilizadas dependendo da aplicação desejada, e o problema da omnisciência lógica é enfatizado. É analisado como o conjunto de fatos conhecidos pelo agente é tratado no momento em que se considera um raciocínio não-monotônico e, por fim, é apresentada uma aplicação em Sistemas Distribuídos, com o enfoque em dois problemas: o problema do "Ataque Coordenado" e o problema dos "Maridos Infiéis". A organização do trabalho é a seguinte:

- No capítulo II, é apresentado o modelo clássico dos mundos possíveis, a "estrutura *kripke*", e os sistemas modais que podem ser usados para capturar as noções de conhecimento e crença.

- No capítulo III, algumas abordagens propostas para capturar a noção de conhecimento e crença são analisadas. O problema da "omnisciência lógica" é definido, sendo considerados inclusive os motivos que levam a este problema.

- No capítulo IV, serão tratadas três abordagens que modelam o raciocínio não-monotônico, sendo feita uma comparação entre as mesmas.

- No capítulo V é apresentado, como um estudo

aplicativo, um modelo geral para um Sistema Distribuído. É definida uma maneira de atribuir conhecimento aos processos e são analisados dois problemas: o problema do "ataque coordenado" e o problema dos "maridos infiéis". No primeiro é analisada a influência da comunicação para atingir o "conhecimento comum" entre os processos do sistema e, no segundo, é analisada a relação existente entre conhecimento, ação e comunicação entre estes processos.

Alguns dos resultados mencionados neste trabalho têm o esboço de sua prova apresentado nos apêndices. Esta organização tem por objetivo captar as principais características de cada abordagem, sem os desvios provocados pela apresentação de detalhes relativos a estas provas.

CAPÍTULO II

FORMALIZAÇÃO DO CONHECIMENTO

Seção 2.1

INTRODUÇÃO

A idéia de formalizar o conhecimento tem por objetivo, entre outros, resolver problemas fundamentais encontrados no campo da ciência da computação relacionados a análise, projeto e compreensão de sistemas complexos cujas partes interagem. Problemas deste tipo são encontrados na Teoria Criptográfica, em Sistemas Distribuídos, Banco de Dados e Robótica.

Este capítulo visa mostrar como a noção de conhecimento pode ser formalizada. Na seção (2.2) será apresentado o modelo clássico dos mundos possíveis; na seção (2.3) será definida a linguagem, que é a lógica proposicional modal para m agentes ($m \geq 1$); segue a seção (2.4), que define um ferramental (a estrutura *Kripke*) que dá à linguagem a semântica desejada. Na seção (2.5) será apresentado o sistema axiomático do conhecimento e na seção (2.6), a crença será formalizada. Na seção (2.7) serão considerados os conceitos de conhecimento comum e conhecimento implícito bem como outras generalizações, que permitirão estabelecer uma hierarquia. Na seção (2.8) será acrescentada a noção do tempo e, por fim, na seção (2.9) será definida um outro ferramental, a estrutura modal, e na seção (2.10) serão dadas algumas conclusões.

Seção 2.2

O MODELO CLÁSSICO

Para formalizar o conhecimento é necessário um modelo semântico. O grande problema para se encontrar este modelo está na definição das propriedades que o conhecimento deve ter, de modo a permitir identificar o que é e o que não é conhecido.

O modelo clássico aceito para esta formalização é o **modelo dos mundos possíveis.** A idéia encontrada neste modelo indica basicamente que, embora o agente se encontre em um estado de conhecimento, o qual determina seu mundo real, é possível associar este agente a um conjunto de estados que poderiam ser reais. Intuitivamente, o estado de conhecimento representa um conjunto de fatos verdadeiros que determinam o mundo em que o agente se encontra. Dá-se o nome de **mundos possíveis** ao conjunto de estados que poderiam ser reais para o agente. O agente não distingue os seus mundos possíveis do mundo real. Desta forma, neste modelo, um agente conhece um fato p se p for verdadeiro em todos os seus mundos possíveis.

Exemplo 2.2.1: Suponha $s_1 = \langle p,q \rangle$ um estado de conhecimento onde os fatos p e q são verdadeiros e A_1 um agente que se encontra neste estado. Suponha $S = \langle s_1, s_2, s_3 \rangle$ o conjunto de mundos possíveis associados a A_1, onde $s_2 = \langle p, \neg q \rangle$ e $s_3 = \langle p,q \rangle$. Logo, pode-se concluir que :

1. A_1 tem o conhecimento de p, pois p é verdadeiro em s_1, s_2 e s_3.
2. A_1 não tem o conhecimento de q, já que q é falso em s_2,

que é um de seus mundos possíveis. ■

Observe que a noção de conhecimento dada só considera fatos conhecidos aqueles que são verdadeiros em todos os mundos possíveis para o agente. Caso em apenas um destes mundos, exista um fato falso, este já não será conhecido pelo agente considerado. Esta noção de conhecimento para um agente pode ser estendida à m agentes, $m \geq 1$ (uma maneira de "ver" estes agentes é associá-los aos processos de um sistema distribuído (ver capítulo V)).

Seção 2.3

A LINGUAGEM

A linguagem considerada para a formalização do conhecimento é a **lógica proposicional modal para m agentes** ($m \geq 1$), definida da seguinte forma:

Definição 2.3.1: (A linguagem L_m)

Dado um conjunto de proposições primitivas $P = \{p,q,...\}$ e um conjunto de m agentes $A = \{1,2,...,m\}$, a **linguagem** L_m é definida como sendo o menor conjunto de fórmulas contendo o conjunto P, fechado sob a negação($\neg$), conjunção($\wedge$) e o operador modal C_i , i = 1,2,...,m, que é operador do conhecimento. A disjunção($\vee$) e a implicação($\supset$) são definidas em função da negação e da conjunção como usualmente.

Exemplo 2.3.1: Suponha p e q $\in$ P. Então, são fórmulas da linguagem : $\neg p$, $(p \wedge q)$, e $C_i p$, onde $C_i p$ indica que o agente i "conhece" o fato p. Já a fórmula $(p \vee q)$ é representada por $\neg(\neg p \wedge \neg q)$ e $(p \supset q)$ por $\neg(p \wedge \neg q)$. ■

Seção 2.4

ESTRUTURA *KRIPKE*

Para dar à linguagem a semântica desejada, define-se uma ferramenta formal, chamada **ESTRUTURA** *KRIPKE*, da seguinte forma:

Definição 2.4.1: (Estrutura *Kripke*)

Uma **estrutura** *Kripke* K é uma n-upla $(S,\pi,\rho_1,\ldots\rho_m)$, onde:

1. S é um conjunto de estados ou mundos possíveis
2. π é uma atribuição de valores verdade às proposições primitivas de P, para cada estado $s \in S$. Logo, $\pi(s,p) \in \{V,F\}$, onde $p \in P$ e $s \in S$.
3. ρ_i é uma relação binária em S onde $i = 1,2,\ldots,m$.

■

Um modelo (mundo) *Kripke* é então definido como um par (K,s), onde K é a estrutura *Kripke* e s ($s \in S$) é um estado que contém uma atribuição de valores verdade às proposições primitivas de P. Se $(s_1,s_2) \in \rho_i$ diz-se que no mundo (K,s_1) o agente i considera (K,s_2) como sendo um mundo possível. Podemos identificar um estado s através de um mundo (K,s) (K é fixo) e dizer que, se $(s_1,s_2) \in \rho_i$, o agente i no mundo s_1 considera s_2 como sendo um mundo possível.

Existe uma classificação, que pode ser imposta a relação ρ_i, que é assim definida:

Definição 2.4.2: Uma relação binária ρ é

1. **Reflexiva** se $(s,s) \in \rho$ para todo $s \in S$
2. **Transitiva** se para todo s_1, s_2, $s_3 \in S$, se $(s_1,s_2) \in \rho$ e

$(s_2,s_3) \in \rho$ então $(s_1,s_3) \in \rho$.

3. **Simétrica** se para todo s_1, $s_2 \in S$, se $(s_1,s_2) \in \rho$ então $(s_2,s_1) \in \rho$

4. **Euclidiana** se para todo s_1, s_2, $s_3 \in S$, se $(s_1,s_2) \in \rho$ e $(s_1,s_3) \in \rho$, então $(s_2,s_3) \in \rho$

5. **Serial** se para todo $s_1 \in S$, existe algum $s_2 \in S$ tal que $(s_1,s_2) \in \rho$

6. De **Equivalência** se ρ é reflexiva, simétrica e transitiva

■

Exemplo 2.4.1: Seja ρ_i uma relação de **equivalência** em S, onde $i = 1,2,\ldots,m$. ρ_i indica que, para o agente i, $(s_1,s_2) \in \rho_i$ se i não for capaz de distinguir o estado s_1 do estado s_2, ou seja s_1 e s_2 pertencem ao conjunto de mundos possíveis para o agente i. ■

Para qualquer modelo (K,s) e uma fórmula $\alpha \in L_m$, define-se quando uma fórmula da linguagem é verdadeira em (K,s), $(K,s) \models \alpha$, indutivamente, quando as seguintes condições são obedecidas:

1. $(K,s) \models p$ (onde $p \in P$) sse $\pi(s,p) = V$
2. $(K,s) \models \alpha \wedge \beta$ sse $(K,s) \models \alpha$ e $(K,s) \models \beta$ (onde α e $\beta \in L_m$)
3. $(K,s) \models \neg\alpha$ sse $(K,s) \not\models \alpha$
4. $(K,s) \models C_i\alpha$ sse $(K,t) \models \alpha$ para todo t tal que $(s,t) \in \rho_i$

O item 4 visa então formalizar o fato do agente i conhecer α no mundo (K,s) quando α é verdadeiro em todos os mundos pertencentes ao seu conjunto de mundos possíveis.

Uma fórmula α é **satisfatível** se existe um modelo (K,s) tal que $(K,s) \models \alpha$. Uma fórmula α é **válida** ($\models \alpha$) se α é

satisfeita em todos os modelos (K,s).

Seção 2.5

O SISTEMA AXIOMÁTICO DO CONHECIMENTO

Antes de definir o sistema axiomático que representa a noção de conhecimento, é necessário apresentar algumas propriedades da relação $\models$, que caracterizam os mundos *Kripke*. São elas:

1. Todas as instâncias de tautologias proposicionais são válidas.
2. Para toda fórmula $\alpha,\beta \in Lm$, $[Ci\alpha \wedge Ci(\alpha \supset \beta)] \supset Ci\beta$ é válida, para $i = 1,2,\ldots,m$.
3. Para toda fórmula $\alpha,\beta \in Lm$, se $\models \alpha$ e $\models \alpha \supset \beta$ então $\models \beta$
4. Para todo $\alpha \in Lm$, se $\models \alpha$ então $\models Ci\alpha$, onde $i = 1,2,\ldots,m$.

Estas propriedades podem ser representadas ao se definir um sistema de axiomas Km, de modo que se possa provar que estes axiomas caracterizam os mundos *Kripke*. Logo, Km consiste de:

1. Dois axiomas:

   A.1 Todas as tautologias do cálculo proposicional

   A.2 $[Ci\alpha \wedge Ci(\alpha \supset \beta)] \supset Ci\beta$, para todo $\alpha,\beta \in Lm$ e $i = 1,2,\ldots,m$

2. Duas regras de inferência:

   R.1. De α e $\alpha \supset \beta$ derive β ("Modus ponens")

   R.2 De α derive $Ci\alpha$ (Caracterização do conhecimento)

O seguinte teorema prova que Km caracteriza os mundos

Kripke:

Teorema 2.5.1 (HALPERN e MOSES, 1985): K_m é uma axiomatização correta e completa para mundos *Kripke*. ■

prova: Apêndice A ■

Para caracterizar o conhecimento, alguns dos axiomas considerados são:

A.3. $C_i\alpha \supset \alpha$, para $i = 1,2,\ldots,m$, que é o **axioma do conhecimento**, indicando que apenas os fatos verdadeiros são conhecidos.

A.4. $C_i\alpha \supset C_iC_i\alpha$, para $i = 1,2,\ldots m$, que é o axioma da **introspecção positiva,** indicando que o agente tem conhecimento do seu conhecimento.

A.5. $\neg C_i\alpha \supset C_i\neg C_i\alpha$, para $i = 1,2,\ldots m$, que é o axioma da **introspecção negativa,** indicando que o agente tem conhecimento da sua falta de conhecimento.

Ao acrescentar A.3 a K_m, o sistema passa a ser conhecido como T_m, acrescentando A.4 a T_m, chama-se o sistema de $S4_m$, e acrescentando A.5 a $S4_m$, o sistema passa a ser chamado de $S5_m$. No caso de $m = 1$, teremos, respectivamente, T, S4, S5 (HUGHES e CRESSWELL, 1968). Quanto aos axiomas A.3, A.4 e A.5, eles só são válidos quando a relação ρ_i é respectivamente reflexiva, transitiva e euclidiana. Dizemos então que um mundo (K,s) é reflexivo (respec. simétrico, transitivo, euclidiano, reflexivo-transitivo, reflexivo-simétrico-transitivo, serial) se todas as relações ρ_i, para $i = 1,2,\ldots,m$, em (K,s) forem reflexivas (respec. simétricas, transitivas, euclidianas, reflexivas-transitivas, reflexivas-simétricas-transitivas, seriais). Com isto, o seguinte teorema pode ser provado:

Teorema 2.5.2 (HALPERN e MOSES, 1985):

1. T_m é uma axiomatização correta e completa para mundos reflexivos
2. $S4_m$ é uma axiomatização correta e completa para mundos reflexivos-tansitivos
3. $S5_m$ é uma axiomatização correta e completa para mundos reflexivos-simétricos-transitivos.

■

prova: Através de argumentos semelhantes àqueles usados na prova do teorema (2.5.1). ■

É válido observar que, ao considerar o sistema S5 (apenas um agente, m = 1), em que ρ é uma relação de equivalência, a estrutura *Kripke* considerada poderá ser "reduzida" apenas ao conjunto S de estados, que representaria o conjunto de mundos possíveis do agente. Cada estado seria então uma atribuição de valores verdade às proposições primitivas. Neste caso, para que $(S,s) \models \alpha$, as condições acima definidas devem ser obedecidas. Podemos então dizer que:

$$(S,s) \models C\alpha \text{ sse } (S,t) \models \alpha \text{ para todo } t \in S$$

Logo, caso exista um mundo $t \in S$ tal que $(S,t) \not\models \alpha$, é fácil ver que $(S,s) \models \neg C\alpha$ para todo $s \in S$.

Ao definir todos estes sistemas, é díficil dizer qual deles melhor traduz a noção de conhecimento. Com certeza, os axiomas de S5 capturam uma boa noção de conhecimento quando se deseja trabalhar com sistemas distribuídos (ver capítulo V).

Seção 2.6

FORMALIZAÇÃO DA CRENÇA

A noção de conhecimento apresentada obriga que apenas fatos verdadeiros podem ser conhecidos. O conceito de crença apareceu com o objetivo de eliminar esta obrigatoriedade, tornando-se bem mais apropriado para a formalização de vários aspectos do raciocínio e para a dedução numa base de conhecimento. Assim, para que a crença possa ser formalizada, o axioma A.3 é retirado do conjunto de axiomas de S5, permitindo que os fatos falsos, apesar de não serem conhecidos, possam ser acreditados. Para evitar problemas de inconsistência na base de conhecimento, deve ser acrescentado a $S5_m$ o seguinte axioma:

A.6 $\neg C_i(falso)$

O sistema resultante da eliminação de A.3 e da adição de A.6 é chamado de $S5_m$-FRACO ou $KD45_m$. Para encontrar um modelo para este sistema, as relações ρ_i devem ser simétricas, transitivas e seriais, e não mais necessariamente reflexivas. Porém, é fácil ver que uma relação binária que é simétrica,transitiva e serial é também reflexiva e este fato é contrário ao tipo de relação que se deseja ter neste novo sistema. Para solucionar este impasse, no caso de um único agente, é bastante considerar uma estrutura em que um estado distinguido descreva o mundo real e um conjunto de estados forme os mundos possíveis. Com isto, a diferença entre o conhecimento e a crença está no fato do mundo real não fazer parte necessariamente do conjunto de mundos possíveis do agente, como exige a noção

de conhecimento. Já no caso de m agentes (m > 1), considera-se a noção intuitiva da relação euclidiana, correspondente ao axioma A.5: para um dado estado s, se ρ_i é euclidiana, então a restrição de ρ_i para $\{t/(s,t) \in \rho_i\}$ é reflexiva, simétrica e transitiva, ou seja, é uma relação de equivalência. Logo, na relação euclidiana, os mundos possíveis formam uma relação de equivalência que não incluem necessariamente o mundo real. Já o fato da relação ser serial irá garantir que o agente sempre terá um conjunto de mundos possíveis. Desta forma, o seguinte teorema pode ser provado:

Teorema 2.6.1 (HALPERN e MOSES, 1985): $S5_m$ FRACO é uma axiomatização correta e completa para mundos euclidianos, transitivos e seriais. ■

prova: Através de argumentos semelhantes àqueles usados na prova do teorema (2.5.1). ■

Com as noções de conhecimento e crença definidas, várias tentativas de estabelecer uma relação entre estas duas noções já foram feitas. Uma destas tentativas estabelece a crença como um conceito básico, definindo o conhecimento em função da mesma, ou seja, o que se tem usualmente é que "o conhecimento é uma crença verdadeira, comprovada". Por exemplo, no caso de uma lei da física que, enquanto for uma conjectura, é considerada uma crença, passando a ser um fato conhecido quando for comprovada teoricamente.

Uma segunda maneira de relacionar estas duas noções seria definir cada uma separadamente e combiná-las do

seguinte modo: cria-se as duas modalidades S5 e S5-fraco, respectivamente, para o conhecimento e a crença, adicionando-se novos axiomas que irão fazer a conexão entre as duas noções. Neste caso, cita-se como desvantagem o fato de não haver uma garantia de que a conexão entre as duas noções seja completamente capturada.

SHOHAM e MOSES (1989) propuseram a seguinte alternativa: define-se o conhecimento e a crença é considerada como uma versão "violável" ou "revogável" deste. Desta forma, as ocorrências do tipo "o agente acredita no fato φ" será traduzida como "o agente tem o conhecimento de φ ou algo específico não usual está ocorrendo". Esta idéia apresenta como vantagem o fato de não ser necessário adicionar nenhuma nova propriedade à lógica do conhecimento, além de sugerir uma conexão estreita entre a noção de crença e o raciocínio não-monotônico.

Definição de crença

A noção de crença é definida em função do conhecimento, estando relacionada também a uma outra fórmula que será uma hipótese. Sendo assim, o agente acredita em uma fórmula φ se ele tem o conhecimento de que φ é verdadeira ou então alguma hipótese φ_{ass} é violada. Logo, como uma tentativa inicial para definir a crença temos:*

Definição 2.6.1: $B'(\varphi, \varphi_{ass}) =_{def} C(\varphi_{ass} \supset \varphi)$ ■

Esta definição captura certas propriedades que sob

determinadas circunstâncias podem não ser aceitas. Por exemplo, $C\neg\varphi_{ass} \supset B'(\varphi, \varphi_{ass})$ é válida, indicando que basta ter o conhecimento de que a hipótese é falsa ($C\neg\varphi_{ass}$ ser verdadeiro) para que acreditemos em φ. Uma segunda definição de crença é então dada acrescentando-se uma condição em que as crenças não sejam limitadas por hipóteses reconhecidamente falsas. O que se deseja é que uma fórmula seja acreditada, relativa a uma hipótese que é conhecidamente falsa, apenas se esta fórmula for conhecida. Seja então a definição:

Definição 2.6.2 : $B(\varphi, \varphi_{ass}) =_{def} C(C\varphi_{ass} \supset \varphi) \wedge (C\neg\varphi_{ass} \supset C\varphi)$ ■

Propriedades da crença

Partindo das definições (2.6.1) e (2.6.2), além das propriedades da crença já vistas, muitas propriedades desejadas aparecem. A seguinte convenção notacional é usada:

1. $B(\varphi, \varphi_{ass})$ será trocado por $B\varphi$ quando a hipótese φ_{ass} puder ser inferida do contexto ou quando a mesma, em particular, não é importante. O mesmo será válido para B'.
2. Quando vários operadores de crença aparecem em uma mesma sentença com as hipóteses omitidas, significa que estas hipóteses são as mesmas para todos estes operadores.

As propriedades abaixo citadas são provadas em (SHOHAM e MOSES, 1989):

1. $B(\varphi, \varphi_{ass}) \equiv C(\varphi \vee (\neg\varphi_{ass} \wedge \neg C\neg\varphi_{ass}))$ é válida: "φ é acreditada sse for conhecido que ou φ é verdadeiro ou a sua

hipótese é violada sem que esta violação seja conhecida".

2. $\neg(B\varphi \wedge B\neg\varphi)$ é válida: "não há crença em expressões contraditórias". Para B', tem-se a validade de : $B'(\varphi,\varphi_{ass}) \wedge B'(\neg\varphi,\varphi_{ass}) \equiv C\neg\varphi_{ass}$

Para relacionar o conhecimento e a crença, temos como fórmulas válidas:

3. $C\varphi \supset B\varphi$, esta também é válida para B' e a validade da regra de inferência "de φ infira $B\varphi$" segue desta propriedade.

4. $BC\varphi \equiv C\varphi$, desta é possível chegar a $BB\varphi \equiv B\varphi$, que é a propriedade da "introspecção positiva".

5. $B\neg C\varphi \equiv \neg C\varphi$, que, da mesma forma, chega a $B\neg B\varphi \equiv \neg B\varphi$, que é a propriedade da "introspecção negativa".

6. $B(\varphi_1 \supset \varphi_2) \supset (B\varphi_1 \supset B\varphi_2)$, que é a propriedade da distributividade.

7. $(BC\varphi \supset \varphi_{ass}) \wedge \varphi_{ass}) \supset \varphi$, indicando que se a hipótese for válida então a crença é verdadeira. Esta também é válida para B'.

Relacionando os operadores B e C, mostra-se por fim estas duas propriedades, que também são válidas:

8. $CB\varphi \equiv B\varphi$

9. $C\neg B\varphi \equiv \neg B\varphi$

O raciocínio não-monotônico

Para capturar a noção da não-monotonicidade, a hipótese presente no operador B (ou B') é vista como uma hipótese na lógica não-monotônica. Ou seja, de acordo com certas hipóteses, certas crenças são adotadas mas, ao

aparecerem novas evidências, é possível descartar algumas delas. Este ponto é tratado quando se examinam as condições sobre as quais um agente pode acreditar, ou até mesmo conhecer, as hipóteses na qual a crença está baseada. Para o operador B', são válidas as seguintes propriedades:

10. $B'(\varphi_{ass}, \varphi_{ass})$

11. $B'(\neg\varphi_{ass}, \varphi_{ass}) \equiv C\neg\varphi_{ass}$

As propriedades (10) e (11) nos afirmam que nós sempre acreditamos na veracidade da nossa hipótese e o único momento em que acreditamos na sua negação é também o único momento em que a crença se torna inconsistente: nós realmente temos o conhecimento de que a hipótese foi violada.

Para o operador B, as propriedades válidas são:

12. $B(\varphi_{ass}, \varphi_{ass}) \equiv \neg C\neg\varphi_{ass}$

13. $B(\neg\varphi_{ass}, \varphi_{ass}) \equiv C\neg\varphi_{ass}$

Ou seja, seu comportamento difere do comportamento de B' apenas quando se trata de acreditar na veracidade da hipótese. Isto é uma inferência não-monotônica já que, enquanto não se tem o conhecimento da falsidade da hipótese, acredita-se que a mesma é verdadeira.

Outras particularidades da noção da crença serão melhor detalhadas no capítulo III, quando for revista a lógica da crença explícita e implícita (LEVESQUE, 1984) Trataremos aqui do conhecimento e suas generalizações.

Seção 2.7

OS ESTADOS DE CONHECIMENTO

Existem situações em que é preciso levar em consideração o estado de conhecimento de um determinado grupo de agentes (salientando que o conhecimento de cada agente deve satisfazer o axioma A.3 da seção (2.5)). Em sistemas distribuídos, por exemplo, há necessidade de uma generalização deste conhecimento individual para o conhecimento em grupo.

O conhecimento Implícito

Uma primeira generalização seria o conhecimento implícito, $CI_g(\varphi)$, onde um grupo g possui este tipo de conhecimento se, e somente, se algum elemento que conhecesse o que cada elemento de g conhece, conseguisse deduzir φ, ou seja, para deduzir φ é preciso ter o conhecimento que está disperso entre os elementos de g.

Exemplo 2.7.1: Suponha um grupo g de dois agentes, a_1 e a_2, e α, β fatos verdadeiros.

Seja $C_1\alpha$ e $C_2(\alpha \supset \beta)$ fatos verdadeiros. Logo, $CI_g(\beta)$ é um fato verdadeiro, já que unindo o conhecimento dos agentes a_1 e a_2 é possível deduzir o fato β. ■

O exemplo acima só é verdadeiro se, dado um estado s, onde $C_1\alpha$ e $C_2(\alpha \supset \beta)$ são fatos verdadeiros ($(K,s) \models C_1\alpha$ sse $(K,t) \models \alpha$ para todo t tal que $(s,t) \in \rho_1$ e $(K,s) \models C_2(\alpha \supset \beta)$ sse $(K,t) \models (\alpha \supset \beta)$ para todo t tal que $(s,t) \in \rho_2$), e dado um conjunto intersecção $\rho = \rho_1 \cap \rho_2$, tenhamos

para todo t tal que (s,t) $\in \rho$, (K,t) $\models \alpha$ e (K,t) $\models (\alpha \supset \beta)$. Logo, podemos deduzir que (K,t) $\models \beta$ e com isto, não teremos apenas $CI_g(\beta)$, mas também $CI_g(\alpha)$ e $CI_g(\alpha \supset \beta)$.

Para a linguagem definida L_m, é bastante acrescentar o operador CI_g, significando o conhecimento implícito. Dada uma estrutura *Kripke* $K = (S, \pi, \rho_1, \ldots, \rho_n)$, o conhecimento implícito é caracterizado de acordo com a seguinte condição:

5. (K,s) $\models CI_g(\varphi)$ sse (K,t) $\models \varphi$ para todo t tal que (s,t) $\in \rho_1 \cap \ldots \cap \rho_n$.

Intuitivamente, φ será verdadeira apenas em mundos pertencentes a intersecção dos mundos possíveis de cada agente. Acrescentando o seguinte axioma e a seguinte regra de inferência aos sistemas axiomáticos já definidos, o conhecimento implícito fica caracterizado por:

A.7.i. $C_i\varphi \supset CI_g(\varphi)$ para i = 1,2,...n

R.3. De $(\varphi_1 \wedge \ldots \wedge \varphi_n) \supset \beta$ derive $(C_1\varphi_1 \wedge \ldots \wedge C_n\varphi_n) \supset CI_g(\beta)$

O conhecimento de "alguém"

Outra generalização é o conhecimento relativo a algum elemento do grupo g sobre um fato φ, $CA_g(\varphi)$. Formalmente tem-se:

$$CA_g(\varphi) = \bigvee_{i \in g} C_i\varphi$$

Exemplo 2.7.2: Considere um grupo g, de três agentes, em que os seguintes fatos são verdadeiros: $C_1\alpha$, $C_2\beta$, $C_3\gamma$. Neste caso, $CA_g(\alpha)$, $CA_g(\beta)$ e $CA_g(\gamma)$ também são verdadeiros.

■

O conhecimento de "todos"

Como terceira generalização, tem-se o conhecimento de todos os elementos do grupo g sobre um dado fato φ. Formalmente, escreve-se:

$$C_{T_G}(\varphi) = \bigwedge_{i \in g} C_i \varphi$$

O $C_{T_G}(\varphi)$ indica que todos os elementos de g conhecem o fato φ. Pode-se definir $(C_{T_G})^{k+1}(\varphi) = (C_{T_G})(C_{T_G})^{k}(\varphi)$, para $k \geq 1$, onde $(C_{T_G})^{1}(\varphi) = C_{T_G}(\varphi)$. Por exemplo , define-se um fato φ como $(C_{T_G})^{2}$**-conhecido** se "todos os elementos de g têm o conhecimento de que todos os elementos de g têm o conhecimento de φ. Da mesma forma, define-se um fato φ como $(C_{T_G})^{k}$**-conhecido** se "todos os elementos de g têm o conhecimento de que todos os elementos de g têm o conhecimento de que...de que todos os elementos de g (k vezes) têm o conhecimento de φ". Formalmente,

$$(C_{T_G})^{K}(\varphi) \equiv \bigwedge C_{ik}(\varphi), \; i \in g$$

Exemplo 2.7.3: Considere apenas dois agentes a_1 e a_2 e $C_{1\alpha}$ e $C_{2\alpha}$ fatos verdadeiros.

Seja $k = 3$, então $(C_{T_G})^{3}(\alpha) \equiv (C_{T_G})(C_{T_G})(C_{T_G})(\alpha) \equiv \bigwedge_{i \in g} C_{i3}(\alpha)$, onde :

$\bigwedge_{i \in g} C_{i3}(\alpha) \equiv C_1C_1C_1\alpha \wedge C_1C_1C_2\alpha \wedge C_1C_2C_1\alpha \wedge C_1C_2C_2\alpha \wedge C_2C_1C_1\alpha \wedge C_2C_1C_2\alpha \wedge C_2C_2C_1\alpha \wedge C_2C_2C_2\alpha \equiv C_1\alpha \wedge C_1C_2\alpha \wedge C_1C_2C_1\alpha \wedge C_1C_2\alpha \wedge C_2C_1\alpha \wedge C_2C_1C_2\alpha \wedge C_2C_1\alpha \wedge C_2\alpha \equiv C_1\alpha \wedge C_1C_2\alpha \wedge C_1C_2C_1\alpha \wedge C_2\alpha \wedge C_2C_1\alpha \wedge C_2C_1C_2\alpha.$

Na linguagem definida L_m, para se obter o conhecimento de "alguém" e o conhecimento de "todos" basta acrescentar os operadores C_{A_G} e C_{T_G}. Quanto a estrutura *Kripke*, para

satisfazer estes dois estados de conhecimento são necessárias as seguintes condições:

6. $(K,s) \models CAg(\varphi)$ sse existe i, i = 1,...,n, tal que para todo t onde $(s,t) \in \rho i$, $(K,t) \models \varphi$.
7. $(K,s) \models CTg(\varphi)$ sse para todo i, i = 1,...,n, tal que para todo t onde $(s,t) \in \rho i$, $(K,t) \models \varphi$.

Nos sistemas axiomáticos já definidos, representa-se os dois estados acrescentando os seguintes axiomas:

A.8. $CAg(\varphi) \equiv C1\varphi \vee \ldots \vee Cn\varphi$

A.9. $CTg(\varphi) \equiv C1\varphi \wedge \ldots \wedge Cn\varphi$

O conhecimento comum

Diz-se que um fato φ é de conhecimento comum, $CCg(\varphi)$, em um grupo g, se, e somente, se entre os elementos de g, φ é verdadeiro e, para $k \geq 1$, φ é $(CTg)^{K}$conhecido. Formalmente,

$$CCg(\varphi) = \varphi \wedge CTg(\varphi) \wedge (CTg)^2(\varphi) \wedge \ldots \wedge (CTg)^n(\varphi) \wedge \ldots$$

Exemplo 2.7.4 : Considere o exemplo anterior. Naquele exemplo, α seria de conhecimento comum sse: α, $CTg(\alpha)$ e $(CTg)^{K}(\alpha)$, para k = 2,3,..., forem fatos verdadeiros.

Levando esta idéia à vida prática, suponha α = "Todos os sinais de trânsito devem ser obedecidos". Para que não ocorram acidentes, não é necessário apenas que todos os elementos do grupo g (da comunidade) conheçam α, mas que todos tenham o conhecimento de que todos tenham o conhecimento de que ...etc.

Para a linguagem de conhecimento comum, além de acrescentar o operador C_G, necessita-se do operador CT_G, do conhecimento de todos.

Para se chegar a estrutura *Kripke* adequada, é necessária a seguinte definição:

Definição 2.7.1: (Relação de acessibilidade)

Diz-se que um estado t é **acessível** de um estado s sempre que existir uma sequência de estados $s_0,\ldots,s_n \in S$, onde $s = s_0$ e $t = s_n$ e, para todo $i = 1,\ldots,n-1$, existir um j com $(s_i,s_{i+1}) \in \rho_j$.

Logo, as condições para formalizar o conhecimento comum são:

1. A condição (7) definida acima e,
2. $(K,s) \models C_G(\varphi)$ sse $(K,t) \models \varphi$ para todo t acessível de s

Quanto ao sistema axiomático para representar o conhecimento comum, acrescentam-se aos sistemas axiomáticos já definidos os seguintes axiomas:

A.9. $CT_G(\varphi) \equiv C_1\varphi \wedge \ldots \wedge C_n\varphi$

A.10. $C_G(\varphi) \supset \varphi$

A.11. $C_G(\varphi) \supset CT_G(C_G(\varphi))$

A.12. $[C_G(\varphi) \wedge C_G(\varphi \supset \beta)] \supset C_G(\beta)$

A.13. $[\varphi \supset CT_G(\varphi)] \supset [\varphi \supset C_G(\varphi)]$

e a seguinte regra de inferência:

R.4. De φ derive $C_G(\varphi)$

Em todas estas generalizações, considerando a linguagem estendida e com o acréscimo dos novos operadores, verifica-se a validade do teorema anterior:

Teorema 2.7.1 (HALPERN e MOSES, 1985):

Para a linguagem de conhecimento comum, conhecimento de "todos", conhecimento de "alguém", conhecimento implícito, incorporando seus axiomas a Km (respec. Tm, S4m e S5m), estes se tornam uma axiomatização correta e completa para mundos *Kripke* (respec. reflexivo, reflexivo-transitivo, reflexivo- simétrico- transitivo). ■

A hierarquia

De acordo com as definições dadas, as noções de conhecimento obedecem a seguinte hierarquia, no sentido das fórmulas abaixo serem válidas:

1. $C_G(\varphi) \supset (CT_G)^k\varphi$ (Todo fato φ que é de conhecimento comum é $(CT_G)^k$-conhecido)
2. $(CT_G)^k\varphi \supset CT_G(\varphi)$ (Todo fato $(CT_G)^k$-conhecido é um fato em que todos em g tem conhecimento).
3. $CT_G(\varphi) \supset CA_G(\varphi)$ (Todo fato φ em que todos em g tem o conhecimento, é um fato conhecido por alguém).
4. $CA_G(\varphi) \supset CI_G(\varphi)$ (Todo fato φ conhecido por alguém, é um fato implicitamente conhecido).
5. $CI_G(\varphi) \supset \varphi$ (Todo fato φ, implicitamente conhecido é um fato verdadeiro).

Vale observar que, se voltarmos ao exemplo (2.7.1), notaremos que $CI_G(\beta) \supset CA_G(\beta)$ (o inverso do item (4)) não é válido, pois em g, não teremos nem $C_1\beta$ nem $C_2\beta$.

Esta hierarquia também pode não ser restrita em determinadas circunstâncias. Basta tomar como exemplo um modelo de computação paralela onde os agentes compartilham uma única memória e o conhecimento de cada um destes

agentes esteja armazenado nesta memória. Neste caso, todas as noções definidas acima deixam de ser distintas (HALPERN e MOSES, 1984a).

Seção 2.8

INCORPORAÇÃO DO TEMPO

Ao acrescentar a noção de tempo à linguagem, é possível tratar de conceitos como aquisição de conhecimento ou esquecimento. Para isto, uma idéia é explicitar a noção de tempo no modelo considerado. Neste modelo (mundos possíveis) a noção de tempo pode ser incorporada ao acrescentar operadores modais e uma relação binária que capture esta noção. Os operadores modais a serem acrescentados são: O e $\hat{\nabla}$ onde $O\varphi$ indica que a fórmula φ é verdadeira no próximo instante de tempo (ou "amanhã") e $\hat{\nabla}\varphi$ indica que φ é eventualmente verdadeira.

Na estrutura *Kripke* considerada, deverá ser então acrescentada uma relação binária T, determinística e serial, de forma que, intuitivamente, $(s,w) \in T$ se w for a descriçao do mundo no próximo instante de tempo após s. Definindo T^* como o fecho reflexivo transitivo de T, formalmente, o que se obtém é:

1. $(K,s) \models O\varphi$ sse $(K,t) \models \varphi$ para um único t tal que $(s,t) \in T$.

2. $(K,s) \models \hat{\nabla}\varphi$ sse $(K,t) \models \varphi$ para algum t tal que $(s,t) \in T^*$.

Desta forma, ao relacionar os operadores de conhecimento e tempo, noções como, por exemplo, o

"não-esquecimento" podem ser capturadas. Para isto, é necessário que o seguinte axioma seja adicionado: $C_i(O\varphi) \supset OC_i\varphi$. Intuitivamente, este axioma indica que se no instante atual o agente tem o conhecimento de que φ será verdadeiro no próximo instante de tempo, então nesse próximo instante o agente terá o conhecimento de φ. Logo, ao ter o conhecimento de φ, o agente adquiriu mais informação e, com isto, serão eliminados do seu conjunto de mundos possíveis aqueles que contrariam o fato de φ ser verdadeiro, ocasionando ("com o passar do tempo") uma redução no número de elementos deste conjunto.

Considere então uma estrutura *Kripke* K e uma relação de equivalência ρ_i. Para que $(K,s) \models C_i(O\varphi)$, devemos ter $(K,w) \models O\varphi$ para todo w tal que $(s,w) \in \rho_i$ e, $(K,w) \models O\varphi$ só ocorre se existir um único estado u tal que $(K,u) \models \varphi$ e $(w,u) \in T$, logo $(s,u) \in \rho_i o T$. Da mesma forma, $(K,s) \models OC_i\varphi)$ só ocorre quando existir um único estado t tal que $(K,t) \models C_i\varphi$ e $(s,t) \in T$. $(K,t) \models C_i\varphi$ só ocorre se $(K,u) \models \varphi$ para todo u tal que $(t,u) \in \rho_i$, isto é, $(s,u) \in T o \rho_i$. Logo, para que o axioma seja válido em K, basta que $T o \rho_i \subseteq \rho_i o T$, ou seja, para três estados s, w, u, se tivermos $(s,w) \in \rho_i$ e $(w,u) \in T$, então existe um estado t tal que $(s,t) \in T$ e $(t,u) \in \rho_i$.

Seção 2.9

A ESTRUTURA MODAL

Em (FAGIN, HALPERN e VARDI, 1984) e (FAGIN e VARDI, 1985) foi definida uma estrutura, chamada **estrutura modal**, que procura descrever explicitamente o que vem a ser um

mundo possível para o agente. A motivação para o estudo dessa estrutura se deve a um problema relacionado às estruturas *Kripke*. Considere o seguinte exemplo:

Exemplo 2.9.1: Seja um estado de conhecimento contendo as seguintes fórmulas: C_1p, $C_2(C_1p)$ e $C_1(C_2(C_1p))$, onde p é uma proposição primitiva. É possível construir estruturas *Kripke* em que estas fórmulas são verdadeiras. O problema se encontra em definir uma estrutura que capture apenas este estado de conhecimento, se é que ela existe. ■

Logo, o problema se encontra na semântica dada as fórmulas que envolvem o conhecimento que, apesar de ser bem definida, a estrutura *Kripke* não deixa claro como deve ser usada para modelar um estado de conhecimento específico.

Para a estrutura modal, o conceito de um mundo possível é utilizado de modo diferente daquele utilizado para estruturas *Kripke* e sua definição obedece a uma certa hierarquia : uma estrutura modal é formada por um conjunto de estruturas modais, onde os mundos são construídos indutivamente em níveis crescentes. Um mundo de nível 0 seria a descrição da realidade, já um mundo de nível 1 seria formado basicamente por um conjunto de mundos de nível 0, um mundo de nível 2, por sua vez, seria formado por um conjunto de mundos de nível 1 e assim por diante.

A noção de conhecimento é modelada por uma estrutura modal, com algumas restrições que são impostas para satisfazer os axiomas de S5, chegando a uma estrutura particular chamada **estrutura de conhecimento**. Intuitivamente, uma estrutura de conhecimento é composta pelos vários níveis, onde o nível 0 corresponde a uma atribuição de valores verdades às proposições primitivas e

o nível k, contém o conjunto de mundos possíveis de ordem k. Considere então o exemplo:

Exemplo 2.9.2: Suponha dois agentes A_1 e A_2 e uma proposição primitiva p, que é verdadeira no nível 0. Para o nível 1, suponha que dois fatos são considerados: A_1 não tem conhecimento se p é verdadeira ou falsa e A_2 tem o conhecimento de que p é verdadeira. Para o nível 2, suponha que A_1 tem o conhecimento de que A_2 tem o conhecimento de que p é verdadeira ou falsa e, com relação a A_2, suponha que A_2 não tem o conhecimento se A_1 tem o conhecimento de p. ■

Neste exemplo, a descrição da "realidade" seria p. Para o nível 1, de acordo com a realidade, um mundo possível para A_1, teria a notação $\langle [p,\neg p] \rangle$, indicando que A_1 não tem o conhecimento da veracidade de p. Para A_2, um mundo possível seria $\langle p \rangle$, indicando que A_2 tem o conhecimento de p. No nível 2 serão gerados os mundos possíveis para cada agente, restrito aos níveis anteriores, indicando o conhecimento de cada agente sobre o seu conhecimento e sobre o conhecimento dos outros agentes. Um mundo possível para A_1, neste nível, poderia ser $w = \langle p, (A_1 \rightarrow \langle [p,\neg p] \rangle, A_2 \rightarrow \langle p \rangle)$. O mundo w indica que p é verdadeiro, e A_1 tem o conhecimento de que não sabe se p é verdadeiro ou falso e também tem o conhecimento de que A_2 conhece p. A notação $A_1 \rightarrow \langle [p,\neg p] \rangle$ representa o fato de A_1 não ter o conhecimento se p é verdadeiro ou falso $(\neg C_1 p)$ e $A_2 \rightarrow \langle p \rangle$ representa o fato de A_2 ter o conhecimento de p $(C_2 p)$. Formalmente, uma estrutura de conhecimento é definida do seguinte modo:

Definição 2.9.1: (estrutura de conhecimento)

Seja um conjunto fixo de proposições primitivas e um conjunto fixo finito de agentes A, onde para cada $i \in A$, existe a modalidade C_i. Defina inicialmente f_0 como uma atribuição de conhecimento de ordem 0 (cada f_0 definido corresponde a uma atribuição de valores verdade às proposições primitivas). Um mundo de ordem 1, é representado por $\langle f_0 \rangle$. Cada agente $A_i \in A$, possui um conjunto de mundos possíveis de ordem 1, $f_1(i) = \langle f_0 \rangle_i$, o qual irá representar o seu conhecimento sobre a atribuição f_0. Assuma indutivamente que mundos de ordem k ($k \geq 1$), $\langle f_0, \ldots, f_{k-1} \rangle$, tenham sido definidos. Defina também W_k como o conjunto de todos os mundos k-ários $\langle f_0, \ldots, f_{k-1} \rangle$. Cada atribuicão de conhecimento de ordem k ($k \geq 1$) é uma função f_k: $A \rightarrowtail 2^{W_k}$, ou seja, cada f_k associa a cada agente um conjunto de mundos k-ários. Logo, o conjunto de mundos possíveis para cada agente, $f_k(i)$, é composto por um conjunto de mundos k-ários restritos ao conhecimento do agente i a cada nível k considerado. Chama-se $\langle f_0, \ldots, f_k \rangle$ um mundo $(k+1)$-ário. A sequencia infinita $\langle f_0, f_1, f_2, \ldots \rangle$ é chamada de estrutura de conhecimento se cada prefixo $\langle f_0, \ldots, f_{k-1} \rangle$ é um mundo k-ário para cada k. ■

Aplicando-se esta definição ao exemplo anterior, teremos:

1. As atribuições de conhecimento de ordem 0 são:

   $f_0 = p$ (descrição da "realidade"), ou

   $f'_0 = \neg p$

   $W_1 = \{p, \neg p\}$

2. Para cada agente $A_i \in A$, um conjunto de mundos possíveis

unários, a partir de f_0, é $f_1(A_i) = \langle f_0 \rangle_i$ definido por:

- $f_1(A_1) = \langle f_0 \rangle_1 = \langle [p, \neg p] \rangle$ (indicando que A_1 não tem o conhecimento da veracidade de p)

- $f_1(A_2) = \langle f_0 \rangle_2 = \langle p \rangle$ (indicando que A_2 tem o conhecimento de que p é verdadeiro)

3. Algumas atribuições de conhecimento de ordem 1, podem ser:

- $f_1 = (A_1 \rightarrow \langle [p, \neg p] \rangle, A_2 \rightarrow \langle p \rangle)$ (corresponde a "realidade", onde A_1 não tem o conhecimento da veracidade de p e A_2 conhece p)

- $f'_1 = (A_1 \rightarrow \langle [p, \neg p] \rangle, A_2 \rightarrow \langle \neg p \rangle)$

- $f''_1 = (A_1 \rightarrow \langle p \rangle, A_2 \rightarrow \langle p \rangle)$

Note que tanto f_1 como f'_1 refletem o fato de A_1 não ter o conhecimento do valor de p. Quanto a A_2, f_1 e f'_1 refletem o fato deste conhecer o valor de p (caso $f_1\langle A_2 \rangle = \langle f_0 \rangle_2$ ou $f_1\langle A_2 \rangle = \langle f'_0 \rangle_2$).

4. Alguns mundos de ordem 2, $\langle f_0, f_1 \rangle$ seriam:

- $w_1 = \langle f_0, f_1 \rangle = \langle p, (A_1 \rightarrow \langle [p, \neg p] \rangle, A_2 \rightarrow \langle p \rangle) \rangle$ (corresponde a "realidade", onde p é verdadeiro, A_1 não tem o conhecimento da veracidade de p e A_2 tem o conhecimento de p)

- $w_2 = \langle f'_0, f'_1 \rangle = \langle \neg p, (A_1 \rightarrow \langle [p, \neg p] \rangle, A_2 \rightarrow \langle \neg p \rangle) \rangle$ (neste mundo, p é falso, A_1 não conhece o valor de p e A_2 conhece o valor de p).

- $w_3 = \langle f_0, f''_1 \rangle = \langle p, (A_1 \rightarrow \langle p \rangle, A_2 \rightarrow \langle p \rangle) \rangle$ (p é verdadeiro e tanto A_1 como A_2 conhecem o valor de p).

5. Para cada agente $A_i \in A$, o conjunto de mundos possíveis de ordem 2 são:

- $f_2(A_1) = \langle f_0, f_1 \rangle_1 = \{w_1, w_2\}$, e

- $f_2(A_2) = \langle f_0, f_1 \rangle_2 = \{w_1, w_3\}$

Note que $f_2(A_1)$ reflete o fato de A_1 saber que A_2 tem o conhecimento do valor de p pois, em w_1, p é verdadeiro e A_2 sabe disso ($A_2 \rightarrow \langle p \rangle$) e, em w_2, p é falso e A_2 sabe disso ($A_2 \rightarrow \langle \neg p \rangle$). Quanto a $f_2(A_2)$, esse reflete o fato de A_2 desconhecer sobre o conhecimento de A_1 sobre p, pois tanto é possível (para A_2), A_1 conhecer o valor de p (em w_3) como A_1 desconhecê-lo (em w_1). ■

Na realidade, o conjunto de mundos gerados a cada nível é composto por mundos que são elementos de um conjunto de mundos possíveis de um agente qualquer $A_i \in A$. Existem então restrições impostas para os mundos $\langle f_0,\ldots,f_k \rangle$, de ordem k + 1, para cada agente i:

1. O mundo $\langle f_0,\ldots,f_{k-1} \rangle$ pertence a um dos conjuntos de mundos possíveis $f_k(A_i)$ ($k \geq 1$). Esta restrição garante que o mundo "real" pertence a todo conjunto de mundos possíveis. No exemplo, $p \in f_1(A_1)$ e $p \in f_2(A_2)$ e, para k = 2, tem-se que $w_1 \in f_2(A_1)$ e $w_1 \in f_2(A_2)$.

2. Se $\langle f'_0,\ldots,f'_{k-1} \rangle \in f_k(A_i)$ e $k > 1$, então $f_{k-1}(A_i) = f'_{k-1}(A_i)$. Esta restrição indica que o agente é introspectivo.

3. $\langle f'_0,\ldots,f'_{k-2} \rangle \in f_{k-1}(A_i)$ sse existir uma atribuição de conhecimento de ordem k-1, f'_{k-1}, tal que $\langle f'_0,\ldots,f'_{k-2},f'_{k-1} \rangle \in f_k(A_i)$, se $k > 1$. Esta restrição indica que um conhecimento de nível mais alto determina (embute) o conhecimento de nível mais baixo.

Dada uma estrutura de conhecimento $f = \langle f_0,f_1,\ldots \rangle$ e uma fórmula α em L_m, para definir o que significa f satisfazer α ($f \models \alpha$), inicialmente define-se:

Definição 2.9.1: (profundidade de uma fórmula α ($d(\alpha)$))

1. $d(p) = 0$ se p é uma proposição primitiva
2. $d(\neg\alpha) = d(\alpha)$

3. $d(\alpha \wedge \beta) = MAX\ (d(\alpha), d(\beta))$

4. $d(C_i\alpha) = 1 + d(\alpha)$

■

Definição 2.9.2: $(\langle f_0,\ldots,f_k\rangle \models \alpha)$

Um mundo $\langle f_0,\ldots f_r\rangle$ de ordem r+1 satisfaz a uma fórmula α, onde $r \geq d(\alpha)$, se:

1. $\langle f_0,\ldots f_r\rangle \models p$, onde p é uma proposição primitiva, se p é verdadeira em f_0.

2. $\langle f_0,\ldots f_r\rangle \models \neg\alpha$ se $\langle f_0,\ldots,f_r\rangle \not\models \alpha$

3. $\langle f_0,\ldots f_r\rangle \models (\alpha \wedge \beta)$ se $\langle f_0,\ldots,f_r\rangle \models \alpha$ e $\langle f_0,\ldots,f_r\rangle \models \beta$

4. $\langle f_0,\ldots f_r\rangle \models C_i\alpha$ se $\langle g_0,\ldots,g_{r-1}\rangle \models \alpha$ para cada $\langle g_0,\ldots,g_{r-1}\rangle \in f_r(i)$. ■

Diz-se que uma estrutura de conhecimento $f = \langle f_0,f_1,\ldots\rangle$ satisfaz a uma fórmula α, $f \models \alpha$, se $\langle f_0,\ldots,f_k\rangle \models \alpha$, onde $k = d(\alpha)$. Uma fórmula α é satisfatível se ela é satisfeita em alguma estrutura de conhecimento e válida se ela é satisfeita em toda estrutura de conhecimento.

Um problema que acontece com a estrutura *Kripke* é que ela não deixa claro que estado de conhecimento corresponde a um estado seu. Esta correspondência é obtida através da relação entre estruturas *Kripke* e estruturas de conhecimento dada no seguinte teorema:

Teorema 2.9.1. (FAGIN, HALPERN e VARDI, 1984): A toda estrutura *Kripke* K e estado s em K, há uma estrutura de conhecimento $f_{k,s}$ correspondente, tal que $(K,s) \models \varphi$ sse $f_{k,s} \models \varphi$, para toda fórmula φ. Da mesma forma, para toda

estrutura de conhecimento f existe uma estrutura *Kripke* K e um estado sf em K tal que f $\models \varphi$ sse (K,sf) $\models \varphi$, para toda fórmula φ. ■

O que o teorema nos diz é que cada nó da estrutura *Kripke* corresponde a uma estrutura de conhecimento em que as mesmas fórmulas são verdadeiras e, para qualquer que seja a estrutura de conhecimento, pode-se construir uma estrutura *Kripke* em que um de seus nós irá satisfazer as mesmas fórmulas da estrutura de conhecimento. Logo, os dois tipos de estruturas se complementam ao modelar o conhecimento, no sentido em que as estruturas de conhecimento modelam cada estado do conhecimento e as estruturas *Kripke* modelam uma coleção de estados de conhecimento.

Seção 2.10
CONCLUSÕES

Diante das várias lógicas modais capazes de modelar o conhecimento, é fácil ver que é possível capturar algumas noções de conhecimento. Isto indica que, apesar destas lógicas permitirem uma boa aproximação para o "raciocínio" necessário a uma base de conhecimento, com o modelo dos mundos possíveis, não é possível formalizar alguns aspectos do raciocínio humano. Isto é devido ao fato dos agentes, neste modelo, serem "perfeitos conhecedores", ou seja, eles sofrem de um problema chamado **omnisciência lógica**, onde o agente conhece todas as fórmulas válidas e todas as consequências lógicas de seu conhecimento. Além disso, sua

adequação para modelar uma base de conhecimento não é muito boa, já que esta é limitada em termos de tempo de computação e espaço de memória. Apesar disto, é um modelo aceito para uma primeira aproximação. Existem algumas abordagens que buscam resolver estes problemas as quais serão analisadas no capítulo seguinte.

Com relação a noção de crença, a proposta de SHOHAM e MOSES (1989) foi definir a crença como sendo o conhecimento relacionado às hipóteses e esta definição está baseada na noção intuitiva de conhecimento e crença. O fato do sistema modal S5 ter sido adotado é justificado por este ser o sistema mais amplamente usado. É discutido então que as propriedades da crença seriam modificadas caso fossem assumidas as "diferentes" lógicas do conhecimento. Neste sentido, alguns problemas ainda permanecem em aberto: um deles seria explorar as propriedades da crença ao assumir outras noções de conhecimento. Outro seria descobrir o que acontece com esta definição quando em noções como a crença comum.

Entre as vantagens e as desvantagens de uma estrutura *Kripke* como um modelo formal de conhecimento, pode-se citar como vantagem o fato de se poder representá-la através de um grafo orientado onde os nós são os mundos possíveis, rotulados pelos valores verdades, de modo que um mundo é possível com relação a outro mundo se existe uma aresta ligando estes mundos. Apesar disto essa estrutura tem seus problemas: com relação a noção de mundos possíveis, por ser considerada primitiva, ela só é bem aplicada quando está clara no contexto, trazendo como exemplo os ambientes distribuídos que não são facilmente modelados por estas

estruturas. Segundo FAGIN, HALPERN e VARDI (1984) uma vantagem das estruturas de conhecimento sobre as estruturas *Kripke* é que as provas de completude são mais "elegantes" e "construtivas", e as provas de decidibilidade são mais fáceis de serem obtidas. Um fato importante é que a satisfatibilidade de uma fórmula em uma estrutura modal dependerá apenas da parte finita da estrutura. As provas de completude e decidibilidade se encontram em (FAGIN e VARDI, 1985).

Através do uso de ferramentas de complexidade computacional, em (HALPERN e MOSES, 1985) são encontrados os detalhes relacionados a complexidade de se determinar a satisfatibilidade de uma fórmula nas lógicas definidas. As conclusões encontradas para este problema dizem que o problema de decidir a satisfatibilidade para S5 e KD45 é NP-completo. Já para K_m, T_m, $S4_m$, para $m \geq 1$, e $S5_m$ e $KD45_m$, para $m \geq 2$, é PSPACE-completo.

CAPÍTULO III

ALGUMAS ABORDAGENS PROPOSTAS PARA FORMALIZAR O CONHECIMENTO

Seção 3.1

INTRODUÇÃO

O modelo clássico, apresentado no capítulo II, apesar de capturar algumas noções de conhecimento, não é bem apropriado para modelar o raciocínio humano. Isto acontece porque, neste modelo, os agentes sofrem do problema da **onisciência lógica**. As dificuldades para tratar este problema se encontram nos seguintes motivos (HALPERN e FAGIN, 1988):

1. A noção de "consciência" do agente relacionada ao seu conhecimento não é levada em consideração. O melhor seria se o agente só pudesse conhecer um fato φ se ele tivesse "consciência" do valor verdade do fato.

2. Faltam recursos computacionais que tenham a capacidade de deduzir todas as consequências lógicas do conhecimento do agente.

3. O agente, em alguns casos, tem o conhecimento incompleto dos fatos que são importantes no processo de dedução.

4. O agente não é capaz de tratar os fatos conhecidos simultaneamente, permitindo que apareçam inconsistências.

Várias abordagens têm sido propostas na literatura para tratar este problema. Cada uma delas tenta modelar os motivos acima apresentados, fazendo com que não seja

possível dizer qual delas é mais eficiente para tratar a onisciência lógica.

Dentre as abordagens que serão analisadas, temos na seção (3.2), a Lógica da Crença Implícita e Explícita de LEVESQUE (1984), na seção (3.3) a lógica da "consciência" de HALPERN e FAGIN (1988), que irá permitir vários agentes e será acrescentado o operador da consciência; na seção (3.4) a abordagem da "sociedade de mentes" onde cada agente irá possuir um conjunto de crenças que podem ser contraditórias; na seção (3.5) a lógica do "raciocínio local" que trata de crenças inconsistentes e, na seção (3.6) a lógica proposicional não-padrão (FAGIN e HALPERN, 1990). Por fim, na seção (3.7) serão dadas algumas conclusões.

Seção 3.2

A LÓGICA DA CRENÇA IMPLÍCITA E EXPLÍCITA (LEVESQUE, 1984)

A lógica de LEVESQUE (1984) formaliza uma idéia de crença mais fraca de modo a evitar não só o problema da onisciência lógica mas também outras desvantagens que surgem ao utilizar a abordagem dos mundos possíveis. Entre estas, encontram-se:

1. O fato de toda sentença válida ser acreditada

2. O fato de que quando se tem duas fórmulas equivalentes, só se acredita em uma delas quando se acredita na outra.

3. O fato de que quando se acredita na fórmula e na sua negação, deve-se acreditar em todas as fórmulas.

Para tratar destes fatos, LEVESQUE (1984) sugere a

formalização de dois tipos de crença: a crença **explícita** e a crença **implícita**, de modo que uma fórmula da linguagem é implicitamente acreditada se ela for consequência lógica de fórmulas explicitamente acreditadas. Logo, o objetivo é, dado uma crença explícita do agente, chegar aos fatos que são implícitos. Como LEVESQUE (1984) diz, o importante é não só o que o agente acredita mas "como o mundo seria idealizado se aquilo que ele acreditasse fosse verdadeiro".

Segundo LEVESQUE (1984), o problema da omnisciência lógica, na abordagem dos mundos possíveis, se encontra no fato do conhecimento e da crença serem caracterizados completamente por um conjunto de mundos possíveis que, além de conter a verdade ou a falsidade dos fatos, possuem todas as tautologias, não sendo possível distinguir uma tautologia de uma verdade contingente. Para fazer esta distinção, a idéia de mundos possíveis é substituída pelo conceito de **situações**:

Uma **situação** é um mundo possível parcial, o que significa que nem todas as fórmulas da linguagem (incluindo as tautologias) têm atribuídas a si um valor verdade. Desta forma, na situação, as fórmulas que não são relevantes para o que o agente realmente acredita (sua crença explícita) não precisam receber um valor verdade. Logo, em uma situação, encontram-se fórmulas verdadeiras e fórmulas falsas, mas serão também encontradas aquelas fórmulas sem nenhum valor lógico, por não serem relevantes à situação.

Exemplo 3.2.1: Considere a situação s significando que "faz frio na região sul". Esta situação irá suportar fatos do

tipo: "As pessoas estão agasalhadas", "A temperatura se encontra abaixo de 20°C", "Algumas casas estão com o aquecedor ligado", etc.. Por outro lado, esta situação pode não suportar fatos do tipo: "Ou as crianças brincam ou estudam", "As lojas estão abertas em horário comercial", "existem ônibus circulando",etc.. Apesar destes últimos serem fatos verdadeiros, eles não são relevantes à situação s. ■

Nessa lógica, a crença explícita é então identificada por um conjunto de situações em substituição aos mundos possíveis. Neste conjunto poderão aparecer situações que suportam tanto a verdade quanto a falsidade de alguma fórmula. Note que, em termos de crença, esta situação representa o fato do agente ter uma visão incoerente do mundo.

A linguagem

Seja o conjunto de proposições primitivas P. A linguagem L_L considerada por LEVESQUE (1984) é formada pelo menor conjunto de fórmulas contendo P e fechado sob a disjunção($\vee$), conjunção($\wedge$) e negação($\neg$) e acrescida dos operadores modais B e L, respectivamente, de crença explícita e implícita, com a restrição de um não poder ocorrer no escopo do outro. A implicação lógica($\supset$) e a equivalência($\equiv$) são representadas em função da disjunção, conjunção e negação.

A estrutura

A estrutura para a crença explícita e implícita é a

n-upla K = (S,$\mathcal{B}$,T,F), onde:

1. S é o conjunto de todas as situações

2. $\mathcal{B}$ é um subconjunto de S, contendo as situações que podem ser reais de acordo com o que é acreditado

3. T e F são funções de P em subconjuntos de S, correspondente as situações que suportam, respectivamente, a verdade e a falsidade de qualquer proposição de P.

As situações podem ser caracterizadas da seguinte forma:

1. Uma situação s é **incoerente** quando ela suporta tanto a verdade quanto a falsidade de uma proposição primitiva $p \in P$ ($s \in T(p) \cap F(p)$). Por outro lado, s é **parcial** se ela não suporta nem a verdade nem a falsidade de $p \in P$ ($s \notin T(p) \cup F(p)$).

2. Uma situação é **completa** quando ela suporta ou a verdade ou a falsidade de toda proposição primitiva p e não é incoerente. Logo, a noção de mundo possível pode ser modelada por esse tipo de situação, desde que toda proposição primitiva cuja verdade (falsidade) é suportada por essa situação seja verdadeira (falsa) no mundo possível.

3. Uma situação completa s é **compatível** com outra situação t caso s se encontre no mesmo conjunto em que t está presente para cada proposição definida.

Dada a estrutura K, resta definir a validade e satisfatibilidade de uma fórmula na linguagem. Para isto, definem-se as relações de suporte $\models_T$ e $\models_F$, onde $s \models_T \alpha$ e $s \models_F \alpha$ indicam respectivamente que a situaçao $s \in S$ suporta a verdade e a falsidade da fórmula α, do seguinte modo:

1. $(K,s) \models_T p$ sse $s \in T(p)$

   $(K,s) \models_F p$ sse $s \in F(p)$

2. $(K,s) \models_T \alpha \vee \beta$ sse $(K,s) \models_T \alpha$ ou $(K,s) \models_T \beta$

   $(K,s) \models_F \alpha \vee \beta$ sse $(K,s) \models_F \alpha$ e $(K,s) \models_F \beta$

3. $(K,s) \models_T \alpha \wedge \beta$ sse $(K,s) \models_T \alpha$ e $(K,s) \models_T \beta$

   $(K,s) \models_F \alpha \wedge \beta$ sse $(K,s) \models_F \alpha$ ou $(K,s) \models_F \beta$

4. $(K,s) \models_T \neg\alpha$ sse $(K,s) \models_F \alpha$

   $(K,s) \models_F \neg\alpha$ sse $(K,s) \models_T \alpha$

5. $(K,s) \models_T B\alpha$ sse $(K,t) \models_T \alpha$ para todo $t \in B$

   $(K,s) \models_F B\alpha$ sse $(K,s) \not\models_T B\alpha$

6. $(K,s) \models_T L\alpha$ sse $(K,t) \models_T \alpha$ para todo $t \in B'$, onde B' é o conjunto de todas as situações completas de S compatíveis com alguma situação em B.

   $(K,s) \models_F L\alpha$ sse $(K,s) \not\models_T L\alpha$

LEVESQUE (1984) define uma fórmula α como sendo válida, representada por $\models \alpha$, se para toda estrutura $K = (S,B,T,F)$ e para toda situação **completa** $s \in S$, $(K,s) \models_T \alpha$ é verdadeira. Uma fórmula α é então satisfeita na situação s se $(K,s) \models_T \alpha$, para alguma estrutura K que tenha s no seu conjunto de situações.

A relação $\models$ tem as seguintes propriedades:

1. Todas as instâncias de tautologias proposicionais são válidas.

2. Todas as tautologias são implicitamente acreditadas e a crença implícita é fechada sob a implicação lógica. O tratamento dado à crença implícita é o mesmo dado ao conhecimento.

3. Toda fórmula que é explicitamente acreditada é implicitamente acreditada ($\models (B\alpha \supset L\alpha)$). A recíproca não é válida. Isto é possível de observar diretamente pelo

próprio conceito de crença implícita e explícita.

Por outro lado, observa-se que:

i. $B\alpha \wedge B(\alpha \supset \beta) \wedge \neg B\beta$ não é válido (a crença explícita não é fechada sob a implicação)

ii. $\neg B(\alpha \vee \neg\beta)$ é satisfatível (uma fórmula válida nem sempre é acreditada explicitamente)

iii. $B\alpha \wedge \neg B(\alpha \wedge (\beta \vee \neg\beta))$ não é válido (nem todo o equivalente lógico de uma crença é acreditado)

iv. $B\alpha \wedge B\neg\alpha \wedge \neg B\beta$ é satisfatível (as crenças são inconsistentes sem que toda a sentença seja acreditada)

A justificativa para a falta de fecho sob a implicação lógica, bem como a crença em fórmulas insatisfatíveis, surgem da possibilidade de se ter situações incoerentes. Da mesma forma, as crenças inconsistentes só são possíveis se toda situação possível de ser acreditada pelo agente for incoerente.

Axiomatização

Para dar uma axiomatização correta e completa a esta lógica, inicialmente é garantido que todas as tautologias estão presentes, validando os três axiomas usuais da lógica e a regra de inferência "Modus Ponens". Para relacionar a crença explícita e implícita, os seguintes axiomas são sugeridos:

AL.1. $L\alpha$ onde α é uma tautologia

AL.2. $B\alpha \supset L\alpha$

AL.3. $(L\alpha \wedge L(\alpha \supset \beta)) \supset L\beta$

Para relacionar a crença explícita com os conectivos lógicos, LEVESQUE (1984) baseia-se em um trabalho anterior

chamado "Lógica de Relevância" (ANDERSON e BELNAP, 1975) que trata do relacionamento entre pares de sentenças, chamado "implicação forte". Assim, aplica-se o conjunto de axiomas para "implicação forte" nesta lógica de relevância:

AL.4. $B(\alpha \wedge \beta) \equiv B(\beta \wedge \alpha)$

$B(\alpha \vee \beta) \equiv B(\beta \vee \alpha)$

AL.5. $B(\alpha \wedge (\beta \wedge \gamma)) \equiv B((\alpha \wedge \beta) \wedge \gamma)$

$B(\alpha \vee (\beta \vee \gamma)) \equiv B((\alpha \vee \beta) \vee \gamma)$

AL.6. $B(\alpha \wedge (\beta \vee \gamma)) \equiv B((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$

$B(\alpha \vee (\beta \wedge \gamma)) \equiv B((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$

AL.7. $B\neg(\alpha \vee \beta) \equiv B(\neg\alpha \wedge \neg\beta)$

$B\neg(\alpha \wedge \beta) \equiv B(\neg\alpha \vee \neg\beta)$

AL.8. $B\neg\neg\alpha \equiv B\alpha$

AL.9. $B\alpha \wedge B\beta \equiv B(\alpha \wedge \beta)$

$B\alpha \vee B\beta \supset B(\alpha \vee \beta)$

Com esta axiomatização, tem-se o seguinte resultado:

Teorema 3.2.1 (LEVESQUE, 1984): (correção e completude)

Uma sentença da linguagem LL é um teorema desta lógica sse ela for uma sentença válida. ■

Aplicações

Esta lógica se adequa às seguintes aplicações:

1. Formalização da crença de outros agentes sem o problema da omnisciência lógica. Como exemplo, apresentam os sistemas de atos de fala, com o objetivo, entre outros, de dar capacidade ao agente para tratar com crenças de outro agente.

2. Representação de conhecimento: considerando uma base de conhecimento formada por um conjunto finito de fórmulas em alguma linguagem proposicional, deseja-se deduzir o valor verdade de alguma proposição. Uma solução possível para este problema consiste em utilizar o que é explicitamente acreditado no lugar de suas implicações. Com isto, um robô, por exemplo, responde baseado nas implicações daquilo que é acreditado.

Em (HALPERN e FAGIN, 1988) são feitas algumas críticas à lógica de LEVESQUE (1984), entre elas, temos:

1. Ao dar a definição de uma fórmula válida, LEVESQUE (1984) considera apenas as situações completas, apesar da relação $\models_T$ ser definida para qualquer situação. Isto garante que toda fórmula proposicional válida continue válida nessa lógica, o que traz inconsistência com a intuição presente no conceito de situações pois, existirão fórmulas válidas nesta lógica, por exemplo, $p \vee \neg p$, tal que $(K,s) \not\models_T p \vee \neg p$ para alguma situação não completa s. Por outro lado, apesar de $p \vee \neg p$ ser uma tautologia proposicional, a relação $\models$ não é bastante clara. Por exemplo, é possível ter uma situação parcial s, onde nem $(K,s) \models_T p$ ($s \notin T(p)$) e nem $(K,s) \models_T \neg p$ ($s \notin F(p)$) e mais uma vez acontecer $(K,s) \not\models_T p \vee \neg p$.

2. LEVESQUE (1984) não trata com crenças aninhadas e só permite um agente. Segundo HALPERN e FAGIN (1988) não é óbvio como estender o modelo de LEVESQUE (1984) para aceitar as crenças aninhadas e mais de um agente.

Em (HALPERN e FAGIN, 1988) observa-se que a falta de conhecimento do agente de fórmulas válidas, bem como a falta do fecho sob a implicação lógica, não é devido a situações incoerentes mas sim ao que é chamado de "consciência" do agente relativa as proposições primitivas. Daí é sugerido o acréscimo à linguagem de um operador modal A ("consciência"), onde Ap significa " o agente tem consciência do valor verdade da proposição primitiva p". Formalmente, Ap é definido como Ap ≡ B(p v ¬p), de forma que Ap é verdadeiro em situações que suportam ou a verdade ou a falsidade de p ou ambas. Assim, já que nem toda fórmula válida é acreditada, nestas condições a crença do agente irá depender da "consciência" do agente com relação a todas as proposições primitivas que aparecem na mesma.

Dada uma fórmula α, defina $A\alpha$ como a conjunção de Ap para toda proposição primitiva p presente em α. Também em (HALPERN e FAGIN, 1988) prova-se que $\models A\varphi \supset B\varphi$. Logo, $B\varphi$ é válido em situações que suportam a verdade ou a falsidade de φ ou ambas, ou seja, em situações em que o agente tem "consciência" de φ. Esta noção de "consciência" na semântica de LEVESQUE (1984) resolve o problema da omnisciência lógica quando este é provocado pela falta de "consciência" do agente relativa a um fato.

O tratamento de crenças aninhadas e de mais de um agente será visto a seguir.

Seção 3.3

A LÓGICA DA CONSCIÊNCIA (HALPERN e FAGIN, 1988)

A lógica da consciência é uma extensão da lógica da crença implícita e explícita, cujo objetivo é suprir algumas deficiências deixadas por LEVESQUE (1984). A primeira delas é o fato de não permitir crenças aninhadas e tratar apenas com um único agente, não sendo possível tratar com noções de conhecimento ou crença relacionadas a um grupo de agentes, e nem mesmo com meta-conhecimento ou crença. Outro problema diz respeito as situações: ao considerar situações incoerentes, LEVESQUE (1984) permite crenças inconsistentes, e com isto, o conjunto de situações possíveis para o agente é um conjunto difícil de ser aceito já que é formado por situações incoerentes.

Nessa lógica será permitido tratar com múltiplos agentes e com crenças aninhadas, tanto na crença explícita quanto na implícita. Essa lógica também não irá permitir situações parciais e incoerentes.

A linguagem

A linguagem é semelhante àquela definida por LEVESQUE, (1984) variando apenas ao permitir uma fórmula especial "*true*" e ao considerar m agentes ($m \geq 1$) e os operadores modais $B_1,\ldots,B_m, L_1,\ldots,L_m$, relativos as crenças explícitas e implícitas destes m agentes. Esta linguagem também permite as crenças aninhadas, de modo que, entre as crenças implícitas e explícitas, uma possa aparecer no escopo da outra.

A estrutura

A estrutura é uma n-upla $K = (S, \pi, A_1, \ldots, A_m, B_1, \ldots, B_n)$, onde:

1. S e π estão definidos na seção (2.4).

2. B_i, $i = 1, \ldots n$, é uma relação binária em S, transitiva, euclidiana e serial.

3. A_i, $i = 1, \ldots, m$, associa cada estado $s \in S$ a um conjunto de proposições primitivas.

Intuitivamente, um par (s,t), com $s, t \in S$, pertence a B_i, se para o agente i no estado s, o estado t é um mundo possível (esta relação é definida como na lógica clássica da crença). Já a função $A_i(s)$ indica o conjunto de proposições primitivas que o agente i tem consciência no estado s.

Como não serão permitidas situações parciais, o estado corresponde a uma situação completa ou mundo possível. As situações parciais podem ser simuladas ao considerar um conjunto ψ de proposições primitivas, e as relações de suporte $\models_T$ e $\models_F$ seriam definidas para este conjunto, de modo a ter cada estado s restrito a uma situação parcial onde apenas as proposições primitivas do conjunto ψ são definidas. Assim, para uma dada proposição primitiva p, teremos:

$T(p) = \{s \in S \;/\; \pi(s,p) = V \text{ (verdadeira)} \text{ e } p \in \psi\}$ e

$F(p) = \{s \in S \;/\; \pi(s,p) = F \text{ (falso)} \text{ e } p \in \psi\}$

Desta forma, para que um estado s suporte a validade de $B_i\varphi$, relativa ao conjunto ψ, é necessário que todos os estados possíveis para o agente i suportem a validade de φ relativa a $\psi \cap A_i(s)$ ($\psi \cap A_i(s)$ é o conjunto das proposições primitivas que i tem consciência no estado s).

A relação $\models$ também é definida nessa lógica e $Bi\varphi$ é verdadeira no estado s, $(K,s) \models Bi\varphi$, se s suporta a verdade de $Bi\varphi$ relativa a $Ai(s)$. Já a crença implícita do agente difere da crença explícita por não considerar a função de consciência, deixando relevante apenas o conjunto de estados possíveis.

As relações de suporte relativas ao conjunto de proposições primitivas $\psi \subseteq P$, $\models_T^\psi$ e $\models_F^\psi$, e a relação que nos dá a noção de validade $\models$ são formalmente definidas por:

1. $(K,s) \models_T^\psi$ "*true*"

   $(K,s) \not\models_F^\psi$ "*true*"

   $(K,s) \models$ "*true*"

2. $(K,s) \models_T^\psi p$ se $\pi(s,p) = V$, onde $p \in \psi$

   $(K,s) \models_F^\psi p$ se $\pi(s,p) = F$, onde $p \in \psi$

   $(K,s) \models p$ se $\pi(s,p) = V$, onde p é uma proposição primitiva.

3. $(K,s) \models_T^\psi \neg\varphi$ sse $(K,s) \models_F^\psi \varphi$

   $(K,s) \models_F^\psi \neg\varphi$ sse $(K,s) \models_T^\psi \varphi$

   $(K,s) \models \neg\varphi$ sse $(K,s) \not\models \varphi$

4. $(K,s) \models_T^\psi \varphi_1 \wedge \varphi_2$ sse $(K,s) \models_T^\psi \varphi_1$ e $(K,s) \models_T^\psi \varphi_2$

   $(K,s) \models_F^\psi \varphi_1 \wedge \varphi_2$ sse $(K,s) \models_F^\psi \varphi_1$ ou $(K,s) \models_F^\psi \varphi_2$

   $(K,s) \models \varphi_1 \wedge \varphi_2$ sse $(K,s) \models \varphi_1$ e $(K,s) \models \varphi_2$

5. $(K,s) \models_T^\psi Bi\varphi$ sse $(K,t) \models_T^{\psi \cap Ai(s)} \varphi$ para todo t tal que $(s,t) \in Bi$

   $(K,s) \models_F^\psi Bi\varphi$ sse $(K,t) \models_F^{\psi \cap Ai(s)} \varphi$ para algum t tal que $(s,t) \in Bi$

   $(K,s) \models Bi\varphi$ sse $(K,s) \models_T^P Bi\varphi$,

6. $(K,s) \models_T^\psi Li\varphi$ sse $(K,t) \models_T^\psi \varphi$ para todo t tal que $(s,t) \in Bi$

$(K,s) \models^{\psi}_{F} Li\varphi$ sse $(K,t) \models^{\psi}_{F} \varphi$ para algum t tal que $(s,t) \in Bi$

$(K,s) \models Li\varphi$ sse $(K,t) \models \varphi$ para todo t tal que $(s,t) \in Bi$

As noções de validade e satisfatibilidade são dadas da maneira usual. Baseando-se nessa definição, HALPERN e FAGIN (1988) provam algumas propriedades:

1. A relação $\models$ é completa, isto é, para cada K, s, α, ou $(K,s) \models \alpha$ ou $(K,s) \models \neg\alpha$

2. Se $\psi \subseteq \psi'$, então: se $(K,s) \models^{\psi'}_{T} \varphi$ então $(K,s) \models^{\psi}_{T} \varphi$

   se $(K,s) \models^{\psi'}_{F} \varphi$ então $(K,s) \models^{\psi}_{F} \varphi$

3. Se $(K,s) \models^{\psi}_{T} \varphi$ ($(K,s) \models^{\psi}_{T} \neg\varphi$) para cada conjunto ψ de proposições primitivas, então $(K,s) \models \varphi$ ($(K,s) \models \neg\varphi$).

4. $\models Bi\varphi \supset Li\varphi$

Além destas, a crença implícita satisfaz o sistema modal S5m-fraco descrito no capítulo II. Para as fórmulas envolvendo crenças aninhadas tem-se: $\models BiLi\varphi \equiv Bi\varphi$.

Outras propriedades encontradas nesta lógica estão também presentes na lógica de LEVESQUE (1984):

1. Todas as tautologias são implicitamente acreditadas e a crença implícita é fechada sob a implicação.

2. Nem toda fórmula válida é explicitamente acreditada e nem todo equivalente lógico de uma crença é acreditado, já que para isto é necessário considerar o conjunto $Ai(s)$.

Pelo fato de não ocorrerem situações incoerentes, pode-se confrontar algumas propriedades dessa lógica com aquela apresentada na seção (3.2): o conjunto de crenças

explícitas é fechado sob a implicação e nenhum agente aceita crenças inconsistentes (Bi(p ∧ ¬p) não é satisfatível). Para LEVESQUE (1984), (K,s) $\models_F B\varphi$ se, e somente se, (K,s) $\not\models_T B\varphi$. Nesta lógica, se (K,s) $\models_F^{\psi} B\varphi$ então (K,t) $\models_F^{\psi \cap Ai(s)} \varphi$, para algum t tal que (s,t) ∈ *Bi*. Assim, é necessário que seja evidente para o agente suportar a falsidade de Biφ.

O relacionamento entre a noção de crença e consciência é dado segundo a intuição definida na seção (3.2), apenas com a extensão para n agentes. Formalmente, teríamos Aip definido por Aip ≡ Bi(p ∨ ¬p), onde p é uma proposição primitiva. Da mesma forma prova-se que, se φ é uma fórmula proposicional válida, Aiφ ⊃ Biφ, onde Aiφ ≡ $\bigwedge_{p \in \varphi}$ Aip. Pode-se relacionar a crença explícita, a crença implícita e a consciência, onde a crença explícita é obtida da combinação entre a crença implícita e a consciência. Por exemplo:

$\models$ Bi(p ∨ q) ≡ (Aip ∧ Lip) ∨ (Aiq ∧ Liq) ∨ (Aip ∧ Aiq ∧ Li(p ∨ q))

Podemos chegar nessa equivalência da seguinte forma:

a) Note que (K,s) $\models_T^P$ Bi(p ∨ q) sse (K,t) $\models_T^{Ai(s)}$ (p ∨ q), ∀ t tal que (s,t) ∈ *Bi*.

Para termos (K,t) $\models_T^{Ai(s)}$ (p ∨ q) dois casos devem ocorrer:

caso 1: p ∈ Ai(s) e π(t,p) = V ou,

caso 2: q ∈ Ai(s) e π(t,q) = V

Seja T_1 = {t / (s,t) ∈ *Bi* e π(t,p) = V} e

T_2 = {t / (s,t) ∈ *Bi* e π(t,q) = V}

Para cada t ∈ $T_1 \cup T_2$, se i tem crença explícita em p ∨ q, temos que:

1. i não tem conciência de q (q ∉ *Ai*(s)). Nesse caso, para

todo t tal que (s,t) ∈ B_i, t ∈ T_1 que implica (K,s) $\models$ Aip ∧ Lip, ou

2. i tem consciência de q. Nesse caso,

2.1. se i não tem consciência de p (p ∉ A_i(s)), então similarmente ao caso 1, (K,s) $\models$ Aiq ∧ Liq.

2.2. se i tem consciência de p, então, dado t tal que (s,t) ∈ B_i, temos que se π(p,t) = F então (K,t) $\vdash_F^{A_i(s)}$p que implica que p ∈ A_i(s) e (K,t) $\vdash_T^{A_i(s)}\neg$p. Por outro lado, se π(p,t) = V então (K,t) $\vdash_T^{A_i(s)}$p e assim, (K,t) $\vdash_T^{A_i(s)}$p ∨ ¬p ⇒ (K,t) $\models$ Bi(p ∨ ¬p) ⇒ (k,t) $\models$ Aip. Similarmente, (K,t) $\models$ Aiq. E se (s,t) ∈ B_i, então, por 1 e 2, t ∈ T_1 ∪ T_2 que implica que (K,t) $\models$ p ou (K,t) $\models$ q, logo, (K,t) $\models$ p ∨ q ⇒ (K,s) $\models$ Li(p ∨ q). Temos então que (K,s) $\models$ Aip ∧ Aiq ∧ Li(p ∨ q).

b) Se (K,s) $\models$ (Aip ∧ Liq) ∨ (Aiq ∧ Liq) ∨ (Aip ∧ Aiq ∧ Li(p ∨ q)) então para todo t tal que (s,t) ∈ B_i, um dos três ítens ocorre:

1. (K,s) $\models$ Aip ∧ Lip, ou seja, p ∈ Ai(s) e ∀ t tal que (s,t) ∈ B_i, (K,t) $\models$ p. Logo, (K,t) $\vdash_T^{A_i(s)}$p ∨ q.

2. (K,s) $\models$ Aiq ∧ Liq é similar ao item 1. Logo, teremos também que (K,t) $\vdash_T^{A_i(s)}$p ∨ q.

3. (K,s) $\models$ Aip ∧ Aiq ∧ Li(p ∨ q), ou seja, p ∈ Ai(s) e q ∈ Ai(s) e ∀ t tal que (s,t) ∈ B_i, (K,t) $\models$ p ∨ q, logo teremos (K,t) $\vdash_T^{A_i(s)}$p ∨ q.

Por 1, 2, 3 temos (K,s) $\models$ Bi(p ∨ q) ■

Na realidade será sempre possível capturar a crença explicita usando uma combinação entre a crença implícita e a consciência, como indica a seguinte proposição:

proposição 3.3.1 (HALPERN e FAGIN, 1988): Dado uma fórmula

φ, é possível obter uma fórmula φ^* tal que $\vdash \varphi \equiv \varphi^*$, onde φ^* é tal que Bi só ocorre nela no contexto de Bi(p v ¬p) e p é uma proposição primitiva. ■

Axiomatização

Para se obter uma axiomatização correta e completa para esta lógica, visto que o operador Li obedece aos axiomas de KD45m, e pelo resultado da proposição (3.3.1), considera-se o seguinte axioma:

A.14. $\varphi \equiv \varphi^*$, onde φ^* é definido como na proposição anterior.

Assim, tem-se o seguinte resultado:

Teorema 3.3.1 (HALPERN e FAGIN, 1988): Acrescentando A.14 aos axiomas de KD45m, obtém-se uma axiomatização correta e completa para a lógica da consciência. ■

prova: Apêndice A ■

Seção 3.4

LÓGICA DE CONSCIÊNCIA GERAL (HALPERN e FAGIN, 1988)

A lógica de consciência geral é uma lógica diferente da lógica da consciência porque, nesse caso, o conjunto $A_i(s)$ não será mais composto apenas por proposições primitivas, mas por fórmulas da linguagem. Uma característica desta lógica é que é possível impor certas limitações ao conjunto $A_i(s)$, cuja importância irá depender da aplicação e, com isto, é possível obter várias interpretações para a noção de consciência. A crença

explícita do agente terá suas limitações em função da noção de consciência e, além disso, esta não será fechada sob a implicação.

A Linguagem

A linguagem é definida da mesma forma que a linguagem anterior acrescentando o operador de consciência A_i, i = 1,...,n, onde $A_i\varphi$ terá várias interpretações dada a noção de consciência, as quais dependerão da aplicação. De modo geral, $A_i\varphi$ pretende significar que "o agente i tem consciência do fato φ".

A Estrutura

A estrutura (*Kripke*) é a n-upla $K = (S, \pi, A_1, \ldots, A_n, B_1, \ldots, B_n)$ onde :

1. S, π e B_i são definidos da mesma forma que na lógica da consciência.

2. $A_i(s)$ é um conjunto de fórmulas, não necessariamente primitivas, onde também "*false*" $\in A_i(s)$ para todo $s \in S$.

Intuitivamente, $A_i(s)$ representa unicamente o conjunto de fórmulas que o agente i "tem consciência" no estado s. Como exemplo, $A_i(s)$ pode ser inconsistente, ou seja, pode conter uma dada fórmula e a sua negação. Também este conjunto pode conter $\psi \vee \varphi$ e não conter $\varphi \vee \psi$, ou conter apenas uma das duas, ou φ ou $\neg\varphi$. Isto permite uma certa flexibilidade na interpretação dada a noção de consciência.

Para dar a noção de validade e satisfatibilidade para as fórmulas da linguagem, define-se apenas a relação $\models$ para fórmulas α da forma $M\varphi$, onde M representa os

operadores modais (não há relações de suporte como definidas na lógica de consciência).

1. $(K,s) \models A_i\varphi$ sse $\varphi \in \mathcal{A}_i(s)$

2. $(K,s) \models L_i\varphi$ sse $(K,t) \models \varphi$ para todo t tal que $(s,t) \in \mathcal{B}_i$

3. $(K,s) \models B_i\varphi$ sse $\varphi \in \mathcal{A}_i(s)$ e $(K,t) \models \varphi$ para todo t tal que $(s,t) \in \mathcal{B}_i$

Nos outros casos a definição é a usual.

Nesta lógica, a crença explícita do agente está diretamente relacionada à sua crença implícita e a sua noção de consciência, onde formalmente temos: $(K,s) \models B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$, ou seja, o agente não terá crenças explícitas de fatos que ele não tenha consciência. Além disso, no caso de se ter um agente "consciente" de todas as fórmulas da linguagem, a sua crença explícita será reduzida a crença implícita. Logo, o operador L_i é considerado o operador clássico da crença e, com relação a crença explícita, da mesma forma que na lógica da consciência, o agente nem sempre acredita explicitamente em todas as tautologias já que, para isto, essas fórmulas devem pertencer ao conjunto $\mathcal{A}_i(s)$. Com relação ao fecho sob a implicação, a fórmula $(B_ip \wedge B_i(p \supset q)) \supset B_iq$, diferente da lógica de consciência não é válida, já que q pode não pertencer a $\mathcal{A}_i(s)$, sendo somente satisfatível.

Como o operador de consciência é um operador essencialmente sintático, existem certas propriedades que podem ser capturadas pela noção de consciência, colocando algumas restrições no conjunto de fórmulas que o agente tem consciência, através da função de consciência. Essas têm sua importância a depender da aplicação. Como exemplo,

algumas das mais comuns são (HALPERN e FAGIN, 1988):

1. Restrição do tipo $\varphi \wedge \psi \in A_i(s)$ sse $\psi \wedge \varphi \in A_i(s)$ torna irrelevante a ordem em que as fórmulas aparecem na conjunção.

2. Restrição que permita o fecho da função de consciência sob as subfórmulas, isto é, se $\varphi \wedge \psi \in A_i(s)$ então φ e $\psi \in A_i(s)$. Esta restrição é utilizada quando se tem uma base de conhecimento cujo valor verdade de suas fórmulas depende do valor verdade das subfórmulas.

3. Restrição em que, para um dado conjunto de proposições primitivas P, o conjunto $A_i(s)$ contivesse apenas fórmulas compostas das proposições primitivas presentes em P. Neste caso, fórmulas como $B_i\varphi \supset B_i(\varphi \vee \psi)$, que são válidas na lógica de consciência, não são mais válidas nesta lógica.

4. Restrição que permita, para dois agentes A_1, A_2, que o agente A_2 não tenha consciência de nenhuma das fórmulas que menciona o agente A_1.

5. Uma restrição que trata da propriedade introspectiva do agente, representada pelo axioma $A_i\varphi \supset A_i(A_i\varphi)$.

6. Uma restrição que permita que o agente determine, em algum tempo especificado ou espaço, se ou não as fórmulas em $A_i(s)$ são deduzidas da informação que ele possui no estado s.

Logo, estas restrições permitem que a função de consciência seja uma ferramenta poderosa para modelar as várias situações. Tratando da crença explícita e implícita, vale salientar que, de acordo com a definição semântica

dada à crença implícita, esta pode ser idêntica a definição de conhecimento dada no modelo clássico dos mundos possíveis (capítulo II).

Axiomatização

Para obter uma axiomatização correta e completa desta lógica, é bastante acrescentar aos axiomas de KD45$_m$, o axioma que define a noção de crença explícita:

A.15. $B_i\varphi \equiv L_i\varphi \wedge A_i\varphi$.

Então,o seguinte teorema é válido:

Teorema 3.4.1 (HALPERN e FAGIN, 1988): Acrescentando o axioma A.15 aos axiomas de KD45m, obtém-se uma axiomatização correta e completa para a lógica da consciência geral. ■

prova: Apêndice A ■

Com relação ao operador B_i, este pode obedecer algumas propriedades que já são obedecidas pelo operador L_i, desde que o operador de consciência A_i seja considerado. Algumas destas são:

1. O fecho sob a implicação $(L_i\varphi \wedge L_i(\varphi \supset \psi) \supset L_i\psi)$ corresponderia a: $(B_i\varphi \wedge B_i(\varphi \supset \psi) \wedge A_i\psi) \supset B_i\psi$.

2. A regra de inferência de φ inferir $L_i\varphi$, corresponderia a: de φ inferir $A_i\varphi \supset B_i\varphi$.

3. Os axiomas de introspecção positiva e negativa ($L_i\varphi \supset L_iL_i\varphi$ e $\neg L_i\varphi \supset L_i\neg L_i\varphi$). Se for assumido que $(s,t) \in \mathcal{B}_i$ implica em $\mathcal{A}_i(s) = \mathcal{A}_i(t)$, captura-se o fato do agente conhecer quais fórmulas ele tem consciência e, desta forma

os axiomas acima corresponderiam a:

$$(B_i\varphi \wedge A_i(B_i\varphi)) \supset B_i(B_i\varphi)$$

$$(\neg B_i\varphi \wedge A_i(\neg B_i\varphi)) \supset B_i(\neg B_i\varphi)$$

Além destas, baseando-se nas restrições que podem ser impostas à função de consciência, a crença explícita pode adquirir novas propriedades que aparecem destas restrições. Como exemplo, considerando o operador da consciência fechado sob as subfórmulas, é fácil provar que a crença explícita do agente é fechada sob a implicação. Existem interpretações naturais da consciência e da crença explícita que não podem ser capturadas simultaneamente com esta lógica (HALPERN e FAGIN, 1988). Como exemplo, basta ter uma interpretação da crença explícita que não pode ser fechada sob a implicação e uma interpretação de consciência que é fechada sob as subfórmulas, este caso não pode ocorrer nesta lógica.

Um exemplo desta lógica é aplicado na análise de sistemas distribuídos (HALPERN e FAGIN, 1988). Neste caso, cada processo do sistema é visto como algum estado local e o sistema como um estado global. Seja s(i) o estado local do processo i no estado global s. Dizer que um processo i tem conhecimento de um fato φ no estado global s, significa que φ é verdadeiro em todos os estados globais s' onde s(i) = s'(i). Logo, os estados globais do sistema correspondem aos mundos possíveis na estrutura *Kripke*. Esta definição faz de B_i uma relação de equivalência nos estados globais de forma que $(s,s') \in B_i$ se, e somente se, s(i) = s'(i).

Para utilizar a noção de consciência, é assumido que o

estado local de cada processo i contém alguma informação e que cada processo i, em um estado global s, está executando algum algoritmo para determinar que fórmulas seguem do fato do estado local de i ser s(i). Estas fórmulas serão as fórmulas que i tem consciência no estado global s, isto é, estas fórmulas pertencem ao conjunto $A_i(s)$. Intuitivamente, só será possível descobrir se $B_i\varphi$ é uma fórmula verdadeira em um estado global s se, ao executar o algoritmo, $\varphi \in A_i(s)$. Os detalhes sobre este assunto não serão vistos, pois não são relevantes neste trabalho.

Seção 3.5

A LÓGICA DO "RACIOCÍNIO" LOCAL (HALPERN e FAGIN, 1988)

A lógica do "raciocínio" local é apresentada com o objetivo de permitir ao agente ter crenças inconsistentes sem fazer uso de situações incoerentes, já que a lógica da consciência geral, apesar de não fazer uso de situações incoerentes, não possui esta propriedade.

A idéia é tratar cada agente como uma "sociedade", composta por estados da mente, contendo o seu próprio conjunto de crenças, os quais podem se contradizer. Esta idéia baseia-se no fato das pessoas pertencerem a grupos que podem não interagir e com isto elas podem possuir crenças inconsistentes. Desta forma, não existirá apenas um conjunto de mundos possíveis para o agente, como na lógica da consciência geral, mas vários conjuntos cada um relacionado a um grupo de crenças diferente.

A linguagem

A linguagem é a mesma definida na seção (3.4), apenas com variações dos operadores modais. Estes são os mesmos definidos no capítulo II para capturar o conhecimento de um grupo de agentes. Dado o operador Bi, onde i = 1,...,n, a fórmula Biφ é interpretada como "o agente i acredita no fato φ em **algum** estado da mente (grupo de crenças)", ou seja, na "sociedade", algum membro acredita em φ. Este tipo de crença é chamada de **crença local** já que ela se refere a um membro da "sociedade". É possível também adicionar operadores modais que capturem noções como a de conhecimento comum ou como a de acreditar em um fato em todos os estados da mente. Nesta abordagem, será adicionado apenas o operador de crença implícita, com o objetivo de ser consistente com os operadores das abordagens anteriores.

O operador de crença implícita nesta lógica é interpretado da mesma forma que o operador de conhecimento implícito definido no capítulo II. Logo, um agente i acredita implicitamente em uma fórmula φ (Liφ) se i puder deduzir φ a partir das informações contidas nos vários estados da mente. Existe uma relação entre a crença local e a crença implícita, onde formalmente um agente i acredita implicitamente em uma fórmula φ, se φ for verdadeira em todo o mundo considerado possível em todos os estados da mente.

A estrutura

A estrutura (*Kripke*) é a n-upla K = (S,π,C_1,...,C_n), onde:

1. S e π são definidos como na seção (3.4)

2. $C_i(s)$ é um conjunto não-vazio de subconjuntos não-vazios de S.

Seja $C_i(s) = \langle T_{i1}, \ldots T_{ik} \rangle$, onde T_{ij}, $j = 1, \ldots, k$, é um subconjunto não vazio de S. Algumas considerações relacionadas a $C_i(s)$ são feitas:

2.1. Para modelar a noção de conhecimento, cada estado $s \in S$, deverá estar presente em todos os subconjuntos $T_{ij} \in C_i(s)$.

2.2. Para a noção de crença, o conjunto de mundos possíveis para o agente i será, em alguns casos T_{i1}, em outros T_{i2}, etc. de modo que cada T_{ij} representará os mundos possíveis para cada membro da "sociedade" do agente i.

Para dar as noções usuais de validade e satisfatibilidade, basta definir a relação $\models$ para fórmulas α da forma $M\varphi$, onde M é um operador modal, pois nos outros casos a definição é a usual:

1. $(K,s) \models B_i\varphi$ sse existe algum $T_{ij} \in C_i(s)$ tal que $(K,t) \models \varphi$ para todo $t \in T_{ij}$.

2. $(K,s) \models L_i\varphi$ sse $(K,t) \models \varphi$ para todo $t \in T_{i1} \cap T_{i2} \cap \ldots \cap T_{ik}$, onde $T_{i1}, \ldots, T_{ik} \in C_i(s)$.

É possível identificar algumas características desta lógica. A primeira delas é o fato da crença explícita não ser fechada sob a implicação: $(B_ip \wedge B_i(p \supset q)) \supset B_iq$ é satisfatível porque cada uma destas crenças individualmente podem estar em diferentes estados, porém não é válida pois, se $C_i(s) = \langle\langle s \rangle, \langle s_1 \rangle\rangle$ e $\pi(p,s) = V$ e $\pi(q,s_1) = F$, e $\pi(q,s)$

$= \pi(p,s_1) = F$ temos que $(K,s) \models Bip$ e $(K,s) \models Bi(p \supset q)$, mas $(K,s) \not\models Biq$. Por esta mesma razão um agente, nesta lógica, pode aceitar crenças inconsistentes ($Bip \wedge Bi\neg p$ é satisfatível), desde que elas pertençam a diferentes estados da mente. Salienta-se também o fato do agente não acreditar em mundos incoerentes, ou seja $Bi(p \wedge \neg p)$ não é válido.

Ao considerar a noção de conhecimento, onde o mundo s em que o agente i se encontra pertence a cada $Tij \in Ci(s)$, as crenças inconsistentes são impossíveis e o axioma $Bi\varphi \supset \varphi$ é válido. Com relação a crença implícita é fácil ver que $Bi\varphi \supset Li\varphi$ é válido e, quando o agente possui crenças inconsistentes, temos que $(Bi\varphi \wedge Bi\neg\varphi) \supset Li(false)$ é válido e, neste caso, não podemos considerar a noção de conhecimento onde $Bi\varphi \supset \varphi$ é válida.

Axiomatização

O sistema KD45m não fornece uma axiomatização correta e completa para esta lógica já que $Li(false)$ é consistente. Logo, são definidos os seguintes axiomas:

A.16. $\neg Bi(false)$

A.17. $Bi\varphi \supset Li\varphi$

Além destes, como a crença local é fechada sob a implicação válida e os agentes acreditam em todas as fórmulas válidas, são sugeridas as seguintes regras de inferência:

R.4 De φ derive $Bi\varphi$

R.5 De $(\varphi \supset \psi)$ derive $(Bi\varphi \supset Bi\psi)$

Daí segue o teorema:

Teorema 3.5.1 (HALPERN e FAGIN, 1988): Os axiomas A.1, A.2 (capítulo II), A.16 e A.17 , e as regras de inferência R.1 (capítulo II), R.4 e R.5 formam uma axiomatização correta e completa para a lógica do raciocínio local. ■

prova: Apêndice A. ■

Outras propriedades podem ser capturadas ao se impor condições ao conjunto $C_i(s)$. Algumas delas são discutidas em (HALPERN e FAGIN, 1988): Por exemplo, para que $\neg L_i(false)$ seja válido, basta garantir que a intersecção dos $T_{ij} \in C_i(s)$ seja não vazia. Um outro exemplo ocorre quando se deseja capturar a noção de conhecimento, onde, para que $B_i\varphi \supset \varphi$ seja válido, devemos garantir que s pertença a cada estado $T_{ij} \in C_i(s)$. Um caso particular ocorre quando o agente, em cada estado da mente, não admite que ele pode estar em outro estado da mente e assim acreditar que é consistente e que tem um "raciocínio perfeito". Com isto, em todas as estrutura que capturam esta propriedade, fórmulas como $B_i(\neg(B_ip \wedge B_i\neg p))$ e $B_i(B_ip \wedge B_i(p \supset q)) \supset B_iq)$ são válidas.

Segundo HALPERN e FAGIN (1988) também é possível adicionar a função de consciência à estrutura *Kripke* desta lógica e capturar um modelo em que o agente não acredite necessariamente em todas as fórmulas válidas já que, para isto, seria preciso que o agente tivesse consciência destas fórmulas. Do mesmo modo, é possível colocar uma restrição em que a consciência do agente fosse fechada sob as subfórmulas sem que sua crença local fosse fechada sob a consequência lógica. Os detalhes de tais extensões não serão analisados neste momento, pois fogem ao escopo deste

trabalho.

Seção 3.6

LÓGICA PROPOSICIONAL NÃO-PADRÃO (FAGIN E HALPERN, 1990)

A abordagem da lógica proposicional não-padrão trata o problema da omnisciência lógica em termos de mundos possíveis não-padrões. Um **mundo padrão** é o mundo até então considerado, onde todas as regras da lógica proposicional (padrão) são verdadeiras. Um **mundo não-padrão** é aquele onde nem todas as regras comuns da lógica proposicional são verdadeiras. Exemplos de mundos não-padrão serão vistos mais tarde.

Esta abordagem considera as implicações da lógica do conhecimento e está baseada em uma **Lógica Proposicional Não-padrão (NPL)** no lugar da lógica proposicional padrão. Não será feita a distinção entre mundos padrões e não-padrões, ou seja, todos estes mundos serão modelos desta lógica. Com isto, o conceito de validade e consequência lógica diz respeito a todos os mundos e o conjunto de mundos possíveis para o agente pode conter mundos padrões ou não-padrões. Desta forma, uma escolha apropriada da NPL traz soluções para o problema da omnisciência lógica.

Mundos possíveis não-padrões

A idéia de propor uma lógica proposicional não-padrão se deve a problemas encontrados em algumas propriedades da lógica proposicional padrão, como por exemplo:

i. A noção de implicação dada a duas fórmulas φ e ψ (φ

$\supset \psi$) é equivalente a $\neg\varphi \vee \psi$, o que não captura a idéia intuitiva da implicação. Exemplo: (p $\supset$ q) $\vee$ (q $\supset$ p) ser válido não é intuitivo visto que p e q podem ser fatos independentes entre si.

ii. Outro problema se relaciona ao fato de uma sentença falsa implicar em qualquer sentença. Exemplo : (p $\wedge \neg$p) $\supset$ q ser válido.

Intuitivamente, para tratar de problemas deste tipo, a NPL propõe que sejam permitidas fórmulas φ e $\neg\varphi$ com seus valores verdade independentes. Logo, o fato de φ ser verdadeiro não irá significar que $\neg\varphi$ seja falso, ou seja, φ pode ser verdadeiro ou falso independente do valor verdade de $\neg\varphi$. Assim, tratando com uma base de conhecimento, seriam considerados verdadeiros todos os fatos que pertencessem a esta base. Desta forma, se dois fatos φ e $\neg\varphi$ fossem encontrados nesta base, então φ e $\neg\varphi$ seriam fatos verdadeiros. Esta abordagem sugere que a todo mundo s esteja relacionado um mundo s^* de modo que uma fórmula $\neg\varphi$ é verdadeira em s se, e somente se, φ não for verdadeira em s^*. Assim s^* dará à semântica as fórmulas negadas de s e quando $s = s^*$, a noção usual de negação é capturada.

A linguagem

A linguagem considerada é a linguagem lógica L_m, definida na seção (2.3).

A estrutura

A estrutura é uma estrutura *Kripke* não-padrão $NK = (S, \pi, \rho_1, \ldots \rho_n, *)$ onde:

1. $(S, \pi, \rho_1, \ldots, \rho_n)$ corresponde a uma estrutura *Kripke*

padrão e,

2. * é uma função unária com domínio em S, onde s^* é resultante da aplicação da função * no mundo s, tal que $s^{**} = s$, para $s \in S$.

A relação $\models$ é definida da mesma forma que para as fórmulas da linguagem na estrutura padrão, exceto para a negação, onde:

$(NK,s) \models \neg\varphi$ sse $(NK,s^*) \not\models \varphi$

Logo, é possível que num mundo s de uma estrutura não-padrão NK, nem φ e nem $\neg\varphi$ sejam fórmulas verdadeiras. Basta que $(NK,s) \not\models \varphi$ e $(NK,s^*) \models \varphi$ $((NK,s^*) \models \varphi$ sse $(NK,s) \not\models \neg\varphi)$. Da mesma forma, pode haver casos em que tanto φ quanto $\neg\varphi$ são fórmulas verdadeiras no mundo s, ou seja, pode-se ter $(NK,s) \models \varphi$ e $(NK,s^*) \not\models \varphi$ $((NK,s^*) \not\models \varphi$ sse $(NK,s) \models \neg\varphi)$. No primeiro caso, o mundo s é definido como um mundo **incompleto** em relação a φ (caso contrário, ele é completo). Já no segundo caso, s é considerado um mundo **incoerente** com relação a φ (caso contrario, ele é coerente). Quando $s = s^*$, o mundo s é um mundo padrão, que é completo e coerente, e neste caso, a semântica da negação é equivalente a definição padrão.

Ao considerar a noção de implicação, onde uma fórmula φ implica logicamente outra fórmula ψ se, e somente se, quando φ é verdadeira então ψ também o é, esta não mais será definida como $\neg\varphi \vee \psi$ em uma estrutura não-padrão NK, devido à semântica não-padrão dada a negação. Basta considerar uma situação onde $(NK,s) \not\models \varphi$, $(NK,s) \not\models \psi$ e $(NK,s^*) \models \varphi$. Considerando NK como uma estrutura *Kripke*

padrão (ignorando a função *) teremos (NK,s) $\models \neg\varphi \vee \psi$. Porém, se NK é uma estrutura não-padrão, teremos (NK,s) $\not\models \neg\varphi \vee \psi$, pois (NK,$s^*$) $\models \varphi$ sse (NK,s) $\not\models \neg\varphi$.

Dado o conjunto de estruturas (não-padrão) e já definida a relação $\models$, a noção de validade e implicação lógica são definidas da maneira usual: se φ é uma fórmula da linguagem L_m, φ é válida em uma estrutura (não-padrão) NK se ela for verdadeira em todos os mundos de toda estrutura (não-padrão) NK. A fórmula φ implica logicamente em outra fórmula ψ em NK se, (NK,s) $\models \varphi$ implica em (NK,s) $\models \psi$ para toda estrutura (não-padrão) NK e todo o mundo s de NK. Em relação a validade, o próximo teorema (teorema (3.6.1)) mostra que nenhuma fórmula é válida para as estruturas não-padrões. A prova deste teorema se dá através de um único contra-exemplo que mostra simultaneamente que nenhuma fórmula de L_m é válida: considere uma estrutura particular NK contendo apenas dois mundos s e t (s,t $\in$ S), tais que s = t^* e t = s^*, onde,

1. $\pi(s,p) = F$ e $\pi(t,p) = V$ para toda proposição primitiva p $\in$ P e,

2. $\rho i = \langle(s,s),(t,t)\rangle$, onde i = 1,...,n.

Por indução, tem-se que, para qualquer fórmula $\varphi \in L_m$, (NK,s) $\not\models \varphi$ e (NK,t) $\models \varphi$. Sendo assim, nenhuma fórmula de L_m é verdadeira em s e, consequentemente, nenhuma fórmula de L_m é válida em estruturas não-padrões NK. Logo, tautologias como $p \vee \neg p$, $(p \supset q) \vee (q \supset p)$, $(\neg p \wedge p) \supset q$, que são válidas na lógica padrão, não são mais válidas nesta lógica.

Teorema 3.6.1 (FAGIN e HALPERN, 1990): Não existe uma fórmula da linguagem padrão L_m que seja válida em estruturas não-padrões. De fato, existe uma estrutura não-padrão NK e um mundo s de NK tal que toda fórmula de L_m é falsa em s e um mundo t de NK tal que toda fórmula de L_m é verdadeira em t. ■

O problema da onisciência lógica, devido a noção de implicação lógica dada à estrutura não-padrão, já não é tão significativo quanto na abordagem padrão. Por exemplo, o conhecimento de fórmulas válidas, que é uma forma de onisciência, é irrelevante já que não existem fórmulas válidas na lógica não-padrão. Também, o conhecimento do agente não é necessariamente fechado sob a implicação, isto é, $Ci\varphi \wedge Ci(\varphi \supset \psi) \supset Ci\varphi$ é satisfatível, mas não é válido.

Implicação forte

Vale observar que, por exemplo, $\neg\neg\varphi$ implica logicamente φ, mas isto não indica que $\neg\neg\varphi \supset \varphi$ é válida como acontece nas estruturas padrões. É definido em (FAGIN e HALPERN, 1990) um conectivo que permite expressar a implicação lógica da mesma forma que o conectivo $\supset$ o faz para estruturas padrões. Além do que, apesar de não se desejar ter fórmulas como $(\neg p \wedge p) \supset q$ sendo válidas, parece bastante razoável aceitar que fórmulas como $\varphi \supset \varphi$ sejam válidas.

É definido um novo conectivo proposicional $\supset_f$, chamado **implicação forte**, que terá o mesmo comportamento do conectivo $\supset$ em estruturas padrões. Logo, uma fórmula $\varphi \supset_f \psi$ será verdadeira se , quando φ é verdadeira, ψ também é

verdadeira. Formalmente,

$(NK,s) \models \varphi \supset_f \psi$ sse (se $(NK,s) \models \varphi$ então $(NK,s) \models \psi$), isto é,

$(NK,s) \models \varphi \supset_f \psi$ sse ou $(NK,s) \not\models \varphi$ ou $(NK,s) \models \psi$

A lógica proposicional não-padrão (NPL)

Define-se L_m^f como o conjunto de fórmulas obtido ao trocar, em L_m, a implicação padrão $\supset$ por $\supset_f$. A **lógica proposicional não-padrão** (NPL) é definida como sendo a parte proposicional de L_m^f e suas interpretações em estruturas não-padrões. A implicação forte $\supset_f$ não pode ser definida em função de $\neg$ e $\vee$ e desta forma, fórmulas como $\varphi \supset_f \varphi$ e $\varphi \supset_f (\varphi \vee \psi)$ são válidas. Existem outras tautologias da lógica padrão que, ao trocar o conectivo $\supset$ por $\supset_f$ não são válidas na lógica não-padrão, por exemplo: $(\varphi \wedge \neg\varphi) \supset \psi$ é válida na lógica proposicional padrão, enquanto que $(\varphi \wedge \neg\varphi) \supset_f \psi$ não é válida na lógica proposicional não-padrão. A implicação forte é então "mais forte" do que a implicação usual visto que, para duas fórmulas padrões φ e ψ, se $\varphi \supset_f \psi$ é válida em estruturas *Kripke* não-padrão, então $\varphi \supset \psi$ é válida em relação a estruturas *Kripke* padrão mas o contrário não acontece.

Com relação a noção de implicação lógica, usando o conectivo $\supset_f$, esta poderá ser expressa como em estruturas padrão:

Proposição 3.6.1 (FAGIN e HALPERN, 1990): Seja φ_1 e φ_2 fórmulas de L_m^f. φ_1 implica logicamente φ_2 em estruturas não-padrão sse $\varphi_1 \supset_f \varphi_2$ é uma fórmula válida em estruturas não-padrão. ■

Na linguagem L_m, uma observação a ser feita diz respeito a afirmação "φ é falsa no estado s". Dado uma estrutura *Kripke* não-padrão NK, não existirá em L_m uma fórmula ψ tal que $(NK,s) \models \psi$ se, e somente se, $(NK,s) \not\models \varphi$. Mesmo uma fórmula da forma $\neg\varphi$, apesar de ser verdadeira, não irá indicar que φ seja falsa num estado s. Mas, ao considerar a linguagem L_m^f, é possível caracterizar quando uma fórmula é falsa num estado s. Acrescentando constantes "*true*" e "*false*" à linguagem, com a semântica usual, a seguinte proposição é verdadeira:

proposição 3.6.2 (FAGIN e HALPERN, 1990): Seja NK uma estrutura não-padrão e s um mundo de NK. Então $(NK,s) \not\models \varphi$ sse $(NK,s) \models \varphi \supset_f$ *false*. ■

Logo, dada a estrutura não-padrão, o único conectivo que possui comportamento diferente da estrutura padrão é a negação ($\neg$). Pode-se então relacionar a lógica proposicional (padrão) e a NPL através de transformações sob fórmulas e estruturas. Na realidade, dada uma estrutura não-padrão NK, é possível obter uma estrutura padrão NK^{*t}, trocando o $*$ de NK pela função identidade. Uma fórmula φ padrão pode ser transformada em uma fórmula não-padrão φ^{nst}, ao trocar todas as subfórmulas da forma $\neg\psi$ por $\psi \supset_f$ *false* e as ocorrências de $\supset$ por $\supset_f$, recursivamente. Da mesma forma a transformação inversa pode ser feita com a troca das ocorrências de $\supset_f$ por $\supset$, desde que a fórmula não-padrão não possua negação. Como resultado destas transformações tem-se que:

Proposição 3.6.3 (FAGIN e HALPERN, 1990): Seja NK uma estrutura não-padrão e s um mundo de NK. Seja φ uma fórmula padrão. Então $(NK,s) \models \varphi^{nst}$ sse $(NK^{st},s) \models \varphi$. ■

Como corolário, temos:

Corolário 3.6.1 (FAGIN e HALPERN, 1990): Seja φ uma fórmula padrão. Então φ é válida em estruturas padrão sse φ^{nst} o for em estruturas não-padrão. ■

Seja φ uma fórmula proposicional. Para provar que φ é válida em L_m^f, inicialmente, através da aplicação de regras de equivalência, transforma-se φ em uma fórmula φ', de modo que em φ' a negação ocorra imediatamente na frente das proposições primitivas. Em seguida, trocam-se todas as ocorrências $\neg p$ de proposições negadas em φ' por uma nova proposição $\bar{p}$, resultando em uma fórmula φ'' livre de negação. Transforma-se, por fim, φ'' em uma fórmula padrão e, pelo corolário (3.6.1), é possível mostrar se φ'' é válida em estruturas não-padrão.

Os seguintes axiomas e regra de inferência dão a axiomatização correta e completa para a NPL:

Defina $\varphi \Leftrightarrow \psi$ uma abreviação para $(\varphi \supset_f \psi) \wedge (\psi \supset_f \varphi)$:

NA.1. Todas as fórmulas φ^{nst} onde φ é uma fórmula válida da lógica proposicional.

NA.2. $\neg\neg\varphi \Leftrightarrow \varphi$

NA.3. $\neg(\varphi \supset_f \psi) \Leftrightarrow ((\neg\psi \supset_f \neg\varphi) \supset_f false)$

NA.4. $\neg(\varphi \wedge \psi) \Leftrightarrow ((\neg\varphi \supset_f false) \supset_f \neg\varphi)$

NR.1. De φ e $\varphi \supset_f \psi$ derive ψ ("modus ponens")

Teorema 3.6.2 (FAGIN e HALPERN, 1990): Os axiomas NA.1,

NA.2, NA.3, NA.4 e a regra de inferência NR.1 dão uma axiomatização correta e completa para a NPL. ■

Para definir um sistema de axiomas que caracterize o conjunto de estruturas *Kripke* não-padrões, basta modificar o sistema de axioma K_m (seção (2.5)) da seguinte forma:

1. Troca-se o "raciocínio" proposicional pelo "raciocínio" proposicional não-padrão e,
2. Troca-se a implicação padrão ($\supset$) nos axiomas e regras de inferência pela implicação forte ($\supset f$).

Com isto, o sistema de axiomas obtido NK_m é formado por:

1. Dois esquemas de axiomas

NA.5. Todas as tautologias do cálculo proposicional não-padrão são válidas

NA.6. $(C_i\varphi \wedge C_i(\varphi \supset f \psi)) \supset f C_i\psi$ (axioma da distribuição)

2. Duas regras de inferência

NR.1. de φ e $\varphi \supset f \psi$ derive ψ ("Modus ponens")

NR.2. De φ derive $C_i\varphi$ (Generalização do conhecimento)

Teorema 3.6.3 (FAGIN e HALPERN, 1990): NK_m é uma axiomatização correta e completa com relação ao conjunto de estruturas não-padrões para as fórmulas na linguagem não-padrão. ■

Seção 3.7

CONCLUSÕES

As lógicas estudadas visam capturar os diferentes

aspectos que causam a falta da omnisciência lógica. Existe a possibilidade de aproveitar a flexibilidade da função de consciência e empregar algumas condições para aplicações particulares. Outro aspecto é tratar de versões quantificadas destas lógicas de modo a atingir uma lógica poderosa o suficiente para descrever situações mais complexas.

Com relação a noção de tempo (seção (2.7)), é possível restringir a função de consciência de modo a permitir que o agente determine o valor verdade das fórmulas, em algum tempo especificado. Para isto, é bastante a seguinte restrição para a função de consciência:

"se $(s,t) \in T$ então $A_i(s) \subseteq A_i(t)$ e para todo $s \in S$ e todas as fórmulas φ, existe algum t com $(s,t) \in T^*$ e $\varphi \in A_i(t)$".

Logo, captura-se a idéia intuitiva de que o conjunto de consciência não decresce com o tempo e de que o agente i tem eventualmente consciência de todas as fórmulas.

Existe ainda uma equivalência entre a semântica dada a Lógica proposicional não-padrão e a Lógica da crença implícita e explícita de LEVESQUE (1984): para cada estrutura NK e um mundo s de NK, pode-se encontrar uma estrutura de LEVESQUE (1984) K e um mundo s' em K tal que para cada fórmula da linguagem φ, $(NK,s) \models \varphi$ sse $(K,s') \models_T \varphi$ e $(NK,s) \models \neg\varphi$ sse $(K,s') \models_F \varphi$. Também, para cada estrutura de LEVESQUE (1984) K e mundo s em K, pode-se encontrar uma estrutura não-padrão NK e um mundo s' em NK tal que para cada fórmula na linguagem φ, $(K,s) \models_T \varphi$ sse $(NK,s') \models \varphi$ e $(K,s) \models_F \varphi$ sse $(NK,s') \models \neg\varphi$ (FAGIN e HALPERN, 1990).

Uma outra relação se encontra na possibilidade de "simular" a lógica da crença implícita e explícita (LEVESQUE, 1984), ao definir a validade e a implicação lógica, em relação aos mundos padrões. Nesta lógica de LEVESQUE (1984), a validade e a consequência lógica são definidas em relação as situações completas, que são mundos padrões, onde todas as regras da lógica padrão são válidas. Dada uma estrutura não padrão NK, um mundo s em NK será padrão se $s = s^*$. Em um mundo padrão, a negação se comporta da maneira usual (NK $\models \neg\varphi$ sse NK $\not\models \varphi$). Formalmente, uma fórmula Lm é **válida em um mundo padrão** se ela for verdadeira em todo o mundo padrão de toda estrutura não padrão.

O problema de decidir a satisfatibilidade de uma fórmula em cada uma das lógicas definidas é provado ser NP-completo nos seguintes casos: Lógica da crença explícita e implícita (LEVESQUE, 1984), Lógica da consciência para o caso de um agente, Lógica do "raciocínio local" para o caso do agente ser consistente e ter um "raciocínio perfeito" (ver seção (3.5)). Ao tratar de fórmulas na lógica de consciência para m agentes ($m \geq 2$), na lógica do "raciocínio local", a lógica proposicional não-padrão e nos casos em que a noção de tempo está presente, o problema de decidir a satisfatibilidade passa a ser PSPACE-completo. O fato da lógica de "consciência geral" não ter sido considerada é justificado por razões que impossibilitam provar sua complexidade. As provas destes resultados são descritas em (HALPERN e FAGIN, 1988) e (HALPERN e FAGIN, 1990) utilizando as técnicas descritas em (HALPERN e MOSES, 1985).

CAPÍTULO IV

A LÓGICA DO CONHECIMENTO E O RACIOCÍNIO NÃO-MONOTÔNICO

Seção 4.1

INTRODUÇÃO

A importância do raciocínio não-monotônico se deve ao fato deste melhor se aproximar do raciocínio humano. Neste sentido, pode-se dizer que as decisões ou conclusões tomadas por um agente estão baseadas em um conjunto finito de fatos que este agente deve conhecer ou acreditar. Desta forma, é possível que algumas conclusões ou decisões, por terem sido obtidas de um conjunto de fatos incompleto devam ser revistas, no momento em que novos fatos são acrescentados neste conjunto.

A noção da não-monotonicidade é então capturada quando, dado um conjunto de axiomas X, inicialmente tenhamos um teorema α derivado de X, mas, quando um novo axioma é adicionado a X, é possível que α não mais seja derivado do novo conjunto X. Nas lógicas clássicas esta noção não está presente pois, nestas lógicas, o fato de aumentar o conjunto de axiomas não irá invalidar os teoremas já derivados do conjunto, existindo apenas a possibilidade de obter novos teoremas.

Um exemplo padrão do raciocínio não-monotônico é o seguinte: A sentença "pássaros voam" não deve indicar o mesmo que "todos os pássaros voam" devido ao fato de existirem exceções, ou seja, existem pássaros que não voam. Esta sentença inicial só será verdadeira quando estivermos

tratando de pássaros típicos, normais, de forma que não exista nenhuma informação que indique o contrário da afirmação.

A idéia é representar a sentença dada por instâncias do tipo "Tipicamente, pássaros voam" ou "Normalmente, pássaros voam". Com isto, para qualquer pássaro em particular, a sentença é verdadeira e, futuramente, caso alguma informação contrária a esse fato se apresente verdadeira para esse pássaro , então conclui-se que esse pássaro não pode voar, ou seja, ele será uma exceção. Logo, considerando A uma hipótese, o que gostaríamos de ter é a seguinte sentença:

"Na ausência de informação do contrário, assuma A"

O problema então se resume a definir esta noção.

Para tratar este tipo de raciocínio vários formalismos foram apresentados, entre eles, podemos citar a **Lógica Default** (REITER, 1980) e a **Circunscrição** (McCARTHY, 1980). No nosso caso, serão analisados alguns formalismos que estão baseados na lógica da crença (ou conhecimento) onde sentenças como "Normalmente, pássaros voam", será traduzida como : "Se x é um pássaro e se é consistente acreditar que x pode voar, então ele voa". Intuitivamente, a noção da não-monotonicidade é atingida em virtude da capacidade que o agente possui de "interferir" no seu conjunto de crenças com o objetivo de inferir sentenças que expressam o que ele não acredita.

Neste capítulo serão revistas três propostas que permitem caracterizar um "estado de conhecimento" de um agente quando, dado uma fórmula α, "α é tudo o que é

acreditado (ou conhecido)" pelo agente. Com isto, a diferença entre estas propostas e as abordagens consideradas anteriormente (capítulo III), se encontra no fato do processo de dedução estar baseado na informação que o agente possui, sem que seja levado em consideração o fato desta informação ser todo o conhecimento (ou crença) do agente. Logo, a idéia é que, dado um conjunto de fórmulas, a crença do agente seja apenas este conjunto. Mas, ao considerar certas hipóteses sobre a capacidade introspectiva do agente, com a condição de que a crença do agente contenha apenas este conjunto de fórmulas, novas crenças poderão surgir. Apresentamos ainda uma quarta proposta que não considera este tipo de raciocínio, mas tem sua importância, pois pode ser usada no estudo dos fundamentos lógicos do TMS, que é um programa que implementa um raciocínio não-monotônico. Nesta proposta o conhecimento do agente estará baseado em dois conjuntos de fatos que irão justificá-lo.

Uma das propostas abordadas é o formalismo apresentado por MOORE (1985), que é uma lógica proposicional contendo suas fórmulas usuais e acrescida do operador de crença. MOORE (1985) chama a sua lógica de **Lógica Autoepistêmica**. Esta lógica foi apresentada como uma nova interpretação da lógica não-monotônica de McDERMOTT e DOYLE (1980).

Uma outra proposta é sugerida por HALPERN e MOSES (1984b). Esta, por sua vez, procura descrever o que o "agente ideal" deverá **conhecer** a partir da informação sobre o mundo em que ele se encontra. Esta abordagem é bem semelhante a anterior sendo que neste caso, a noção de conhecimento é usada enquanto no caso anterior, MOORE

(1985) baseia-se na noção de crença.

Uma característica destes dois formalismos é que os fatos em que o agente acredita (ou tem conhecimento) estão presentes em um conjunto e a noção de implicação lógica especifica que fatos o agente pode inferir deste conjunto.

Uma terceira proposta que será revista é a lógica definida por LEVESQUE (1990), a qual pode ser vista como uma versão objetiva da lógica autoepistêmica de MOORE (1985). Nesta lógica, o operador B de crença é utilizado explicitamente e com isto, a noção de implicação lógica é definida da maneira usual. Logo, nesta abordagem, além do operador usual de crença B, é introduzido o novo operador modal O ("only knowing"), onde Oα irá significar "α é tudo o que é acreditado".

Um exemplo que traduz o tipo de raciocínio proposto por estes três formalismos é o seguinte:

Exemplo 4.1.1 (LEVESQUE, 1990): Sejam as premissas:

1. Tweety é um pássaro
2. Se é consistente acreditar que um pássaro pode voar, então ele voa

Logo, é possível chegar a conclusão de que

3. Tweety voa

Mas, ao aplicar a segunda premissa a Tweety, é obtido que:

4. É consistente com o que é acreditado que Tweety voa

O fato (4) é justificado por não haver outras crenças sobre Tweety além de (1) e (2), ou seja:

5. Isto é "tudo o que é acreditado" sobre Tweety

Logo, o fato (4) pode ser derivado de (1) e (5), onde

(5) nos diz que (1) e (2) é "tudo o que é acreditado" sobre Tweety. Além disso, ainda é possível ter o fato (3) justificado por (1), (2) e (5). ■

Neste exemplo, o fato que irá produzir inferências não-monotônicas é o fato (2). Observa-se que os únicos pássaros que não podem voar são aqueles que podem ser inferidos a não voar. Este tipo de raciocínio é chamado de **raciocínio autoepistêmico**, já que ele apresenta um raciocínio a respeito da própria crença (ou conhecimento) do agente. Ele é não-monotônico porque é "sensível ao contexto", ou seja, a noção dada no fato (2) de "ser consistente acreditar", se refere apenas ao conjunto de fatos (premissas) em questão. Logo, ao acrescentar um novo fato a este conjunto de premissas, por exemplo: "Tweety não voa", não mais teremos a conclusão apresentada no fato (3).

A última proposta que apresentaremos é a **lógica do conhecimento envolvido** definida por RAO e FOO (1987a, 1987b). Nesta lógica, o conhecimento do agente sobre um fato φ dependerá de um conjunto de premissas de φ que o agente deverá ter conhecimento e de certas exceções que o agente deverá desconhecer. Na realidade, para que o agente conheça φ, ele deverá "justificar" o seu conhecimento e esta justificativa é determinada tanto por fórmulas que são conhecidas quanto por fórmulas que o agente desconhece.

Este capítulo será exposto da seguinte forma: na seção (4.2) será tratada a lógica autoepistêmica de MOORE (1985), na seção (4.3) será vista a lógica proposta por HALPERN e MOSES (1984b). A seção (4.4) trata da proposta de LEVESQUE (1990). A seção (4.5) mostra a relação entre os três

formalismos apresentados, os quais traduzem o raciocínio autoepistêmico. A seção (4.6) apresenta a lógica do conhecimento envolvido definida por RAO e FOO (1987a, 1987b) e, por fim, na seção (4.7), serão dadas algumas conclusões. Convém lembrar que, nos quatro formalismos apresentados, o agente é considerado um "agente ideal", com capacidade introspectiva e raciocínio proposicional perfeitos.

Seção 4.2

A LÓGICA AUTOEPISTÊMICA DE MOORE (MOORE, 1985)

MOORE (1985), em sua lógica autoepistêmica, define uma teoria como sendo um conjunto de fatos que representariam as crenças do agente. Desta forma, é possível inferir novas crenças a partir destes fatos, desde que os mesmos sejam consistentes com a teoria em questão. Por tratar com a noção de crença, o sistema modal considerado é o S5-fraco.

A linguagem

A linguagem considerada é a lógica proposicional acrescida de um operador modal autoepistêmico. Nesta lógica, por estar baseada na noção de crença, constará o operador modal B para significar "é acreditado"*. MOORE (1985) limita a sua linguagem ao caso proposicional, justificando que ocorrem problemas com o significado do quantificador no escopo do operador autoepistêmico B.

MOORE (1985) utiliza a letra L mas, como já associamos o perador B a crença, este será utilizado.

Para modelar as crenças do agente, é definido um conjunto de fórmulas da linguagem como o conjunto de crenças do agente.

Definição 4.2.1: (Teoria Autoepistêmica)

Uma **Teoria Autoepistêmica** T é um conjunto de fórmulas (que representam as crenças do agente). ■

Semântica

Uma fórmula Bα será considerada verdadeira para o agente se, e somente se, α pertencer ao seu conjunto de crenças (T). Dado então uma teoria autoepistêmica T, as crenças do agente serão determinadas por:

1. Constantes proposicionais que são verdadeiras (em um mundo considerado) e,

2. fórmulas que o agente acredita

Para então formalizar esta noção semântica, MOORE (1985) nos dá as seguintes definições:

Definição 4.2.2: (Interpretação proposicional e Modelo proposicional)

Seja T uma teoria autoepistêmica. Uma **Interpretação Proposicional** de T é uma atribuição de valores verdade às fórmulas da linguagem em T. Esta atribuição deve ser consistente com a recursão usual para a lógica proposicional, e com qualquer atribuição arbitrária de valores verdade para as constantes proposicionais e para fórmulas do tipo Bα. Um **Modelo Proposicional** de T é uma interpretação proposicional de T em que todas as fórmulas de T são verdadeiras. ■

Neste caso, vale observar que as fórmulas do tipo Bα são tratadas como constantes proposicionais (pode-se ter, por exemplo, Bα ∈ T e α ∉ T).

Definição 4.2.3: (Interpretação autoepistêmica e Modelo autoepistêmico)

Seja T uma teoria autoepistêmica. Uma **Interpretação Autoepistêmica de T** é uma interpretação proposicional de T, onde para toda fórmula α, Bα é verdadeira se, e somente se, α pertencer a T. Um **Modelo Autoepistêmico de T** será então uma interpretação autoepistêmica de T em que todas as fórmulas de T são verdadeiras. ■

MOORE (1985) observa que a própria teoria T determina o valor verdade das fórmulas do tipo Bα, independente da atribuição de valores verdade às constantes proposicionais de T que ocorrem em α. Com isto, para cada atribuição de valores verdade para estas constantes, existe exatamente uma interpretação autoepistêmica correspondente.

Exemplo 4.2.1 (MOORE, 1985): Seja T uma teoria autoepistêmica. Suponha que T determine as crenças do agente em um dado mundo w. O mundo w conterá uma atribuição de valores verdade às constantes proposicionais de T. Qualquer fórmula do tipo Bα será verdadeira para o agente se, e somente se, α ∈ T. Desta forma, o agente e o mundo w determinam uma interpretação autoepistêmica de T. Caso todas as crenças do agente forem verdadeiras em w, esta interpretação será um modelo autoepistêmico de T. ■

MOORE (1985) define então a noção de inferência da

seguinte forma: "O problema da inferência é um problema em que se deseja determinar que conjunto de crenças (teoremas) o agente adotaria com base em suas premissas iniciais (axiomas)". Logo, as crenças deste agente devem ser "corretas" e "semanticamente completas" em relação às premissas iniciais. A "correção" e a "completude" são assim definidas:

Definição 4.2.4: (correção e completude)

Seja T uma teoria autoepistêmica e Pc um conjunto inicial de premissas. T é **correto** em relação a Pc se, e somente se, toda interpretação autoepistêmica de T, na qual todas as fórmulas de Pc são verdadeiras, é um modelo autoepistêmico de T. T é, por sua vez, semanticamente **completa** se, e somente se, T contém toda fórmula que é verdadeira em todo modelo autoepistêmico de T. ■

A noção de correção captura a condição de que todas as crenças do agente sejam verdadeiras quando todas as suas premissas são verdadeiras. Dada uma interpretação autoepistêmica de T que é determinada pelo que é verdadeiro em um certo mundo w, todas as crenças do agente serão verdadeiras em w se todas as suas premissas também forem verdadeiras. Caso seja considerada uma interpretação autoepistêmica de T em que todas as fórmulas de Pc sejam verdadeiras, mas algumas fórmulas de T sejam falsas, então, apesar das premissas do agente serem verdadeiras em w, algumas de suas crenças não o serão.

Em relação à noção de "completude", temos que se α é verdadeira em todo o modelo autoepistêmico de T, então ela

deve ser verdadeira quando todas as crenças do agente são verdadeiras. Com isto, o agente é capaz de inferir α. Caso α seja falsa em algum modelo autoepistêmico de crenças do agente, então este poderá ser um dos mundos possíveis para o agente e com isto o agente não irá acreditar em α.

Sintaxe

Resta então definir a sintaxe para as teorias autoepistêmicas de modo a satisfazer as condições de correção e completude. Segundo MOORE (1985), quando se trata de lógicas não-monotônicas, o procedimento usual de definir regras de inferência que se aplicam aos axiomas falha, pois, com estas regras, não é possível inferir fatos em um determinado estágio, e futuramente quando algum outro fato for inferido, este passe a invalidar fatos que tenham sido inferidos anteriormente. O que se faz então é especificar as condições que definem o conjunto de crenças (T) que o agente deve possuir:

1. Se $\alpha_1,\ldots,\alpha_n \in T$ e $\alpha_1,\ldots,\alpha_n \models \beta$ então $\beta \in T$ (onde, $\models$ significa consequência tautológica).
2. Se $\alpha \in T$ então $B\alpha \in T$
3. Se $\alpha \notin T$ então $\neg B\alpha \in T$

Este conjunto T passa a ser chamado de **Teoria Autoepistêmica Estável**. Para que este conjunto possa então ser consistente, mais duas condições são acrescentadas:

4. se $B\alpha \in T$, então $\alpha \in T$
5. se $\neg B\alpha \in T$, então $\alpha \notin T$

Ao obedecer estas condições (2 a 5), T se torna correta e samanticamente completa em relação a fórmulas da forma Bα e ¬Bα: Bα ∈ T sse α ∈ T e ¬Bα ∈ T sse α ∉ T. Logo, todo o modelo proposicional de uma teoria autoepistêmica estável é um modelo autoepistêmico. Outros resultados são :

1. O valor verdade de qualquer fórmula de uma teoria autoepistêmica estável dependerá apenas do valor verdade das fórmulas da teoria que não possuem o operador de crença (fórmulas objetivas).

2. Como as fórmulas objetivas de uma teoria autoepistêmica estável determinam se as fórmulas da teoria são verdadeiras, então estas fórmulas determinam todas as fórmulas que estão contidas na teoria.

Um outro resultado apresentado mostra que se uma teoria autoepistêmica T é estável então T é semanticamente completa, muito embora não garanta sua "correção" em relação ao seu conjunto inicial de premissas. Na realidade, ainda é possível para o agente acreditar em proposições que não estão presentes nas suas premissas iniciais. Com isto, é adicionada uma restrição impondo que as únicas proposições que o agente deve acreditar são aquelas que pertencem ao seu conjunto de premissas e aquelas necessárias as condições de estabilidade. Logo, uma teoria autoepistêmica T passa a incluir apenas as conseqüências lógicas de Pc ∪ {Bα / α ∈ T} ∪ {¬Bα / α ∉ T}. MOORE (1985) define então que esta teoria T está **fundamentada** em um conjunto de premissas Pc, isto é, α ∈ T se, e somente se, α for consequência lógica de Pc e de fórmulas do tipo Bα e ¬Bα, respectivamente, quando α ∈ T e quando α ∉ T. Outra

conclusão sua diz que os únicos conjuntos de crenças que são válidos para o agente, dado Pc como premissas, contém apenas as extensões de Pc que estão fundamentadas em Pc e são estáveis. Estes conjuntos são chamados de **Expansões Estáveis** de Pc.

É possível, para um conjunto de premissas, encontrar mais de uma expansão estável ou até mesmo nenhuma expansão estável. Considere os exemplos (MOORE, 1985):

Exemplo 4.2.2: Seja Pc = {¬Bα ⊃ β, ¬Bβ ⊃ α}. Em qualquer teoria autoepistêmica estável que inclua Pc, se α não estiver na teoria, β estará e se β não estiver na teoria, ¬Bβ estará e consequentemente α estará. Mas, para que a teoria esteja fundamentada nestas premissas, α só estará na teoria se ela for consequência lógica de Pc e de fórmulas do tipo Bφ se φ ∈ T e ¬Bφ se φ ∉ T. Logo, ¬Bβ deverá estar na teoria. Assim, α será consequência lógica de ¬Bβ ⊃ α e Pc e ¬Bβ, não sendo possível incluir β na teoria. Da mesma forma, se β estiver na teoria, também não há meios de incluir α. Logo,poderemos ter uma expansão estável de Pc contendo α e da mesma forma poderemos ter uma expansão estável de Pc contendo β, mas não poderemos ter uma expansão estável de Pc contendo α e β. ■

Exemplo 4.2.3: Se considerarmos Pc = {¬Bα ⊃ α}, Pc não terá nenhuma expansão estável. Suponha que T é uma teoria autoepistêmica estável baseada em Pc. Se α ∈ T, pela estabilidade Bα ∈ T. Porém, pela definição de fundamentação, α está em T por consequência lógica de fórmulas de Pc e fórmulas do tipo Bφ e ¬Bφ que estariam em

T. Portanto ¬Bα deveria pertencer a T, mas ¬Bα ∈ T se, e somente se, α ∉ T, o que seria um absurdo. Por outro lado, se α ∉ T, então ¬Bα ∈ T e por consequência lógica de Pc, α ∈ T o que seria também uma contradição. Logo, não existe uma teoria autoepistêmica estável fundamentada em Pc. ■

MOORE (1985) levanta o problema de como "ver" sua Lógica autoepistêmica como uma lógica, ou seja, ao definir o conjunto Pc de premissas como axiomas, e determinar o que seriam os teoremas. Dois casos são considerados:

1. Pc possui apenas uma expansão estável. Neste caso, o conjunto de teoremas seria esta expansão estável.
2. Pc possui várias expansões estáveis ou nenhuma expansão estável. Neste caso, de acordo com a semântica desejada, a idéia é considerar para o agente conjuntos alternativos de teoremas ou até mesmo nenhum conjunto de teoremas de Pc. Uma outra idéia é definir este conjunto de teoremas como a intersecção de todas as fórmulas da linguagem, com todas as expansões estáveis de Pc. Segundo MOORE (1985), estas duas sugestões são bem razoáveis, muito embora a segunda possua uma interpretação diferente da primeira, pois não considera o ponto de vista do agente, representando o que seria conhecido quando se conhece apenas as premissas do agente.

Seção 4.3

A ABORDAGEM PROPOSTA POR HALPERN E MOSES (1984b)

HALPERN e MOSES (1984b) apresentam uma abordagem muito semelhante a lógica Autoepistêmica de MOORE, onde a idéia

de se ter "α é tudo o que é conhecido" é também caracterizada, mas neste caso o sistema modal obedecido é o S5.

A motivação para definir esta abordagem se relaciona ao problema da comunicação em sistemas distribuídos (ver capítulo V). O objetivo era caracterizar o estado de conhecimento do processo em um dado ponto no tempo, visto que a comunicação modifica o estado de conhecimento dos processos do sistema. O estado de conhecimento de cada processo representaria "tudo o que é conhecido" pelo processo. Outro problema levantado foi o fato de certas fórmulas não caracterizarem um "estado de conhecimento". Por exemplo, se um processo i conhece apenas a fórmula $Cp \vee Cq$ ($Cp \vee Cq$ é "tudo o que é conhecido"), i não terá o conhecimento de p já que i conhece p ou conhece q e, da mesma forma, i não terá o conhecimento de q. Mas este "estado de conhecimento" é inconsistente pois i não pode conhecer p ou conhecer q sem ter o conhecimento isolado de um deles.

A proposta é então introduzir não só uma abordagem semelhante a lógica autoepistêmica, caracterizando um "estado de conhecimento" em que "α é tudo o que é conhecido" pelo agente, mas é apresentada também esta caracterização com base na semântica dos mundos possíveis.

A linguagem

A linguagem considerada também será a lógica proposicional acrescida, neste caso, do operador de conhecimento C.

A semântica

Como neste caso o objetivo é modelar o conhecimento do agente, define-se então o **estado de conhecimento** T do agente como o conjunto de fórmulas conhecidas pelo agente, de forma que, uma fórmula da linguagem α pertence a T se, e somente se, Cα for uma fórmula verdadeira. Para que T venha a ser um estado de conhecimento, as propriedades que definem as condições de estabilidade a serem satisfeitas por T neste caso serão:

1. T contém todas as instâncias de tautologias proposicionais
2. Se $\alpha \in T$ e $\alpha \supset \beta \in T$ então $\beta \in T$
3. $\alpha \in T$ sse $C\alpha \in T$
4. $\alpha \notin T$ sse $\neg C\alpha \in T$
5. T é (proposicionalmente) consistente

Ao obedecer a estas condições, T é chamado de **conjunto estável** de fórmulas. A condição (5) implica que o conjunto de todas as fórmulas da linguagem, por ser inconsistente, não é um estado de conhecimento. Da mesma forma que na lógica autoepistêmica de MOORE (1985), é provado em (HALPERN e MOSES, 1984b) que um conjunto estável é determinado unicamente pelas fórmulas proposicionais que ele contém.

Dado uma fórmula α, o estado de conhecimento do agente quando "α é tudo o que é conhecido", deveria ser em algum sentido minimal entre os estados de conhecimento contendo α. Mas isto não acontece. É provado em (HALPERN e MOSES, 1984b) que não é possível comparar dois estados de

conhecimento com respeito a inclusão, ou seja, nenhum conjunto estável inclui propriamente outro conjunto estável. Como não é possível comparar dois estados de conhecimento, uma alternativa para a minimalidade seria um conjunto estável contendo α cujo **subconjunto proposicional** é mínimo em relação à inclusão. O problema surge porque não é para todas as fórmulas α que este conjunto pode ser encontrado. Um exemplo seria: seja α uma fórmula do tipo $Cp \vee Cq$. Qualquer conjunto estável contendo α, deverá conter ou p ou q. Logo, poderá existir um conjunto estável T_p contendo α e p sem conter q, e um conjunto estável T_q contendo α e q sem conter p. O que ocorre é que a intersecção entre os subconjuntos proposicionais de T_p e T_q não conterá nem p nem q, nos levando a crer que não existe um conjunto estável T contendo α, cujo subconjunto proposicional esteja contido tanto no subconjunto proposicional de T_p como no subconjunto proposicional de T_q.

Definição 4.3.1 : (Fórmula honesta)

Uma fórmula α é **honesta** se, e somente se, existe um conjunto estável contendo α cujo subconjunto proposicional é mínimo. Este conjunto é chamado de T^{α}. Em caso contrário ela é **desonesta.** ■

Exemplo 4.3.1: Como visto acima, $\alpha = Cp \vee Cq$ é uma fórmula desonesta. ■

Logo, para uma fórmula honesta α, T^{α} descreve o estado de conhecimento do agente quando "α é tudo o que é

conhecido".

Uma outra maneira na qual HALPERN e MOSES (1984b) caracterizam este estado de conhecimento é utilizando a semântica dos mundos possíveis, cujo modelo *Kripke* obedece ao sistema modal S5. Como se trata de um único agente, a estrutura *Kripke* K considerada será um conjunto S não-vazio de estados (que são atribuições de valores verdade às proposições primitivas da linguagem). Estes estados serão os mundos possíveis para o agente (seção (2.5)).

Dado então uma estrutura *Kripke* K, define-se o conjunto de fórmulas conhecidas em K, C(K), da seguinte forma:

$$C(K) = \{\alpha \mid (K,s) \models \alpha \text{ para todo } s \in S\}.$$

Como resultado, a seguinte proposição relaciona um conjunto estável a uma estrutura *Kripke*:

Proposição 4.3.1 (HALPERN e MOSES, 1984b): Todo conjunto estável T determina um modelo *Kripke* K_T no qual $T = C(K_T)$. Além disso, se a linguagem possuir apenas um número finito de proposições primitivas, então K_T é o único modelo *Kripke* com esta propriedade ($K_T = \{s \mid s$ satisfaz as fórmulas proposicionais de T$\}$) ■

Como um corolário desta proposição, tem-se:

Corolário 4.3.1 (HALPERN e MOSES, 1984b): Os conjuntos estáveis são fechados sob a consequência lógica S5. ■

Com este corolário, a condição 1 de estabilidade pode

ser trocada por :

1'. T contém todas as instâncias de tautologias de S5.

Dado um modelo *Kripke* K e uma fórmula α, resta definir quando esta estrutura satisfaz a "α é tudo o que é conhecido" pelo agente. Intuitivamente, o conjunto de mundos possíveis desse modelo será maximal entre os conjuntos de mundos possíveis das estruturas *Kripke* que satisfazem Cα. Se considerarmos duas estruturas *Kripke* K_1 e K_2, se $K_2 \subset K_1$ então o conhecimento do agente em K_1 é "menor" do que o conhecimento do agente em K_2, desde que o conjunto de mundos possíveis K_1 para o agente é maior. Logo, o modelo apropriado é aquele em que o agente conhece o "mínimo" entre todos aqueles nos quais o agente conhece α. Este modelo será então aquele em que o conjunto de mundos possíveis é o maior de todos nos quais Cα é verdadeira. Sendo K^{α} o modelo desejado, diz-se que α é **honesta**$_k$ se $\alpha \in C(K^{\alpha})$. K^{α} é o modelo resultante da união dos conjuntos de mundos possíveis de todos os modelos em que Cα é verdadeira.

Exemplo 4.3.3 (HALPERN e MOSES, 1984b): Seja $\alpha = Cp \vee Cq$ e seja $K_p = \{s \mid \pi(s,p) = V\}$ e $K_q = \{s \mid \pi(s,q) = V\}$. Logo, K_p e K_q satisfazem a Cα. Ao formar o conjunto contendo todos os mundos possíveis $K_p \cup K_q$, seria obtido o modelo de "maior ignorância", onde os únicos fatos conhecidos seriam as tautologias de S5. Com isto, $\alpha \notin C(K^{\alpha})$, logo α não é uma fórmula honesta$_k$. ■

O exemplo (4.3.3) ilustra um resultado mais geral que

é: as noções de fórmulas que são honesta$_k$ e honesta$_T$ coincidem, bem como C(K$^\alpha$) e T$^\alpha$, para uma fórmula honesta α.

HALPERN e MOSES (1984b) também argumentam que o problema de honestidade de uma fórmula é decidível.

Seção 4.4

"ALL I KNOW" : A ABORDAGEM PROPOSTA POR LEVESQUE (LEVESQUE, 1990)

Na lógica autoepistêmica, MOORE (1985) define uma interpretação subjetiva, onde a presença de uma fórmula em uma teoria autoepistêmica indica que a mesma é acreditada pelo agente "dono" da teoria. LEVESQUE (1990), nesta sua abordagem, considera uma lógica objetivamente interpretada e, neste caso, para que a fórmula seja acreditada, esta não é colocada na teoria como um axioma mas utiliza-se um operador de crença explicitamente. Três vantagens são citadas ao considerar esta interpretação objetiva:

1. Ela expressa tudo o que é expresso em uma interpretação subjetiva
2. Com esta interpretação, os conceitos usuais da lógica clássica não precisam ser modificados (um destes conceitos, por exemplo, seria a noção de consistência e de consequência lógica).
3. Como as restrições na crença são representadas por fórmulas da linguagem, é possível determinar as condições sobre as quais o sistema "define" suas crenças e como estas crenças se "enquadram" no mundo.

A idéia é estabelecer uma relação entre a noção de crença e a não-monotonicidade, onde as propriedades do

"raciocínio" do agente são expressas como fórmulas de uma linguagem modal contendo quantificadores e com semântica voltada para a noção dos mundos possíveis.

A Linguagem

A linguagem considerada, chamada de OL, é uma linguagem modal, de 1ª ordem com igualdade, cujo operadores modais são B e O, e seus únicos símbolos funcionais formam um conjunto infinito contável de símbolos funcionais 0-ários (símbolos constantes). Essas constantes são chamadas de nomes padrões ou parâmetros. Se n é um nome padrão e α uma fórmula cuja única variável livre é x, α^x_n denota a fórmula resultante da troca de toda ocorrência livre de x em α pelo nome padrão n.

As fórmulas são formadas da maneira usual, sem restrições no escopo dos quantificadores e operadores modais. Podemos destacar algumas classes de fórmulas:

1. fórmulas **objetivas**: são aquelas em que não ocorrem os operadores B e O.

2. fórmulas **subjetivas**: são aquelas onde todos os símbolos não-lógicos ocorrem no escopo de B ou O.

3. fórmulas **básicas**: são aquelas que não contém o operador O

4. fórmulas **atômicas**: são as letras e predicados proposicionais (exceto o de igualdade) aplicados aos nomes padrões.

Os nomes padrões terão uma função semelhante a das constantes no universo de Herbrand, onde os termos sem variáveis seriam gerados a partir destes nomes padrões.

Nesta abordagem, como na lógica autoepistêmica, a

noção de crença apresentada satisfaz ao sistema modal S5-fraco. A semântica das fórmulas é apresentada, inicialmente, para as fórmulas básicas, acrescentando posteriormente para as fórmulas em que ocorrem o operador O. Para dar a semântica das fórmulas básicas, são considerados dois fatores independentes:

1. Quais fórmulas atômicas são verdadeiras e,
2. Quais fórmulas são acreditadas.

Com estes dois fatores, as outras fórmulas básicas terão seus valores fixados pelas regras recursivas usuais.

Para as fórmulas atômicas, será utilizada uma função w das fórmulas atômicas em {0,1}, chamada de **atribuição** onde, para cada fórmula ϕ, $w(\phi) = 1$ indica que ϕ é verdadeira. Dada uma fórmula α, para determinar se $B\alpha$ é uma fórmula verdadeira, utiliza-se um conjunto de atribuições W, de modo que $B\alpha$ é verdadeira se, e somente se, para todo $w \in W$, α é verdadeiro em relação a w. Logo, pode-se relacionar cada atribuição w a um mundo possível e cada conjunto de atribuições W como um conjunto de mundos possíveis. Assim, formalmente, teremos a seguinte definição:

Dada uma fórmula básica α, um conjunto de atribuições W e uma atribuição w, a noção de satisfatibilidade é assim definida:

1. Se ϕ é atômica, $(W,w) \models \phi$ sse $w(\phi) = 1$
2. Se ni e nj são nomes padrões, $(W,w) \models (ni = nj)$ sse ni e nj são idênticos (são os mesmos símbolos)
3. $(W,w) \models \neg\alpha$ sse $(W,w) \not\models \alpha$
4. $(W,w) \models (\alpha \wedge \beta)$ sse $(W,w) \models \alpha$ e $(W,w) \models \beta$
5. $(W,w) \models \exists x\alpha$ sse para algum nome padrão n, $(W,w) \models \alpha^x_n$
6. $(W,w) \models B\alpha$ sse para todo $w' \in W$, $(W,w') \models \alpha$

De acordo com os ítens (5) e (6), é válido citar a distinção entre B∃xα e ∃xBα, onde no segundo caso, para todo w' ∈ W, deve-se ter (W,w') ⊢ $\alpha^{\overline{x}}_{n}$, e com isto, o mesmo n é usado para cada w' ∈ W. Já no primeiro caso, para cada w' ∈ W, é possível ter um n diferente. Dado então um conjunto Γ de fórmulas básicas, (W,w) ⊢ Γ se, e somente se, para todo α ∈ Γ, (W,w) ⊢ α.

Axiomatização

LEVESQUE (1990) determina para a parte básica de OL, os seguintes axiomas e regras de inferência:

axiomas:

AO.1. (ni = ni) ∧ (ni ≠ nj) onde ni e nj são nomes padrões distintos

AO.2. Todas as instâncias de teoremas da lógica de primeira ordem

AO.3. Bα onde α é uma instância de um teorema da lógica de primeira ordem

AO.4. B(α ⊃ β) ⊃ (Bα ⊃ Bβ), onde α e β são fórmulas quaisquer

AO.5. (∀xBα) ⊃ (B∀xα) (generalização universal)

AO.6. (σ ⊃ Bσ) onde σ é uma fórmula subjetiva

Regras de inferência:

RO.1. De α e α ⊃ β derive β ("modus ponens")

RO.2. De $\alpha^{\overline{x}}_{n_1},\ldots,\alpha^{\overline{x}}_{n_k}$ derive ∀xα se ni varia sobre todos os nomes em α e sobre pelo menos um que não ocorre em α.

"Only Knowing"

Ao acrescentar o operador O, o que se pretende é

capturar a seguinte idéia: As crenças são fórmulas que são verdadeiras em todos os mundos possíveis . Logo, dado uma nova fórmula objetiva, acreditar nesta nova fórmula significa reduzir o conjunto de mundos possíveis, deixando apenas aqueles em que a nova crença é verdadeira. Com isto, a fórmula Oα ("α é tudo o que é acreditado") indica que o mínimo possível é acreditado dado que α é acreditado, ou seja, o conjunto de mundos possíveis é tão maior quanto forem os mundos em que a fórmula α for verdadeira. A regra semântica para Oα é a seguinte:

7. (W,w) $\models$ Oα sse (W,w) $\models$ Bα e, para todo w', se (W,w') $\models$ α então w' ∈ W

Pela regra semântica (6) acima, poderemos reescrever a regra (7) da seguinte forma:

7'. (W,w) $\models$ Oα sse para todo w', w' ∈ W sse (W,w') $\models$ α

Intuitivamente, por (6) todos os mundos possíveis satisfazem α e por (7'), apenas os mundos possíveis satisfazem α.

É fácil ver que a noção de satisfatibilidade para fórmulas do tipo Bα e Oα depende apenas do W em questão. Logo, podemos alterar a notação (W,w) $\models$ Oα ((W,w) $\models$ Bα) por W $\models$ Oα (W $\models$ Bα). No caso de fórmulas objetivas, estas dependem unicamente do w considerado e, neste caso, para uma fórmula objetiva ψ, poderemos escrever w $\models$ ψ no lugar da notação (W,w) $\models$ ψ.

Validade e satisfatibilidade

Para dar a noção de validade e satisfatibilidade, LEVESQUE (1990) inicialmente aborda o seguinte fato: os

estados de crença (autoepistêmicos) têm sido tratados em termos de fórmulas que são acreditadas, muito embora sejam modelados através de conjuntos de atribuições. É também observado que existem menos estados de crença caracterizados por fórmulas acreditadas do que estados de crença caracterizados por conjuntos de atribuições. Com isto, muitos conjuntos de atribuições diferentes representam o mesmo conjunto de crenças. Estes conjuntos são chamados de conjuntos **equivalentes**. A idéia então é definir uma semântica que não permita distinguir dois conjuntos equivalentes. É sugerido que seja escolhido um conjunto para representar cada classe de equivalência (formada por conjuntos equivalentes), que seria o maior conjunto da classe. LEVESQUE (1990) mostra então que este conjunto é único e sempre é possível chegar a ele sem alterar o valor verdade de qualquer fórmula básica. A este conjunto será dado o nome de conjunto **maximal autoepistêmico**.

Definição 4.4.1: (conjunto maximal autoepistêmico)

Um **conjunto maximal autoepistêmico** é aquele que não possui um superconjunto que é equivalente a ele. Logo, um conjunto maximal autoepistêmico pode ser interpretado como o conjunto de todas as atribuições que satisfazem a toda crença básica de algum conjunto de atribuições. ■

Ao considerar o operador O, LEVESQUE (1990) define a noção de validade e satisfatibilidade para OL relacionada aos conjuntos maximais autoepistêmicos de atribuições:

Definição 4.4.2: (conjunto satisfatível)

Um conjunto Γ é **satisfatível** se, e somente se, existir um conjunto maximal autoepistêmico W e uma atribuição w tal que $(W,w) \models \varphi$, para toda fórmula φ em Γ. Γ **implica** uma fórmula α ($\Gamma \models \alpha$) se, e somente se, $\Gamma \cup \{\neg\alpha\}$ é insatisfatível. Finalmente, α é **válida** sse for implicada pelo conjunto vazio. ■

Exemplo 4.4.1 (LEVESQUE, 1990): Seja ψ uma fórmula atômica da lógica de primeira ordem e W o conjunto de todas as atribuições w tais que $w(\psi) = 1$. Logo, $(W,w) \models O\psi$, pois $(W,w) \models B\psi$ e se $(W,w') \models \psi$ então $w' \in W$. Diz-se então que "ψ é tudo o que é acreditado". Note que W é um conjunto maximal autoepistêmico de atribuições pois, se $w' \notin W$ ($w'(\psi) = 0$) fosse adicionado a W, ter-se-ia $(W,w) \not\models B\psi$ e portanto $(W,w) \not\models O\psi$. ■

A noção de conjuntos maximais autoepistêmicos, apesar de não ser levada em consideração quando se determina a satisfatibilidade de uma fórmula básica, é importante quando se deseja determinar a satisfatibilidade de qualquer fórmula subjetiva. LEVESQUE (1990) mostra isto através do seguinte exemplo:

Exemplo 4.4.2: Seja W e ψ definidos como no exemplo anterior. Seja Σ o seguinte conjunto:

$\Sigma = \{B\alpha \mid \alpha$ é básica e $W \models B\alpha\} \cup \{\neg B\alpha \mid \alpha$ é básica e $W \models \neg B\alpha\}$

O problema é determinar se $\Sigma \models O\psi$, ou seja, se $\Sigma \cup \{\neg O\psi\}$ é insatisfatível. A resposta seria negativa se

conjuntos não-maximais autoepistêmicos fossem permitidos: basta considerar um $w \in W$ e um $W^* = W - \{w\}$. W^* é então um conjunto equivalente a W, mas é um conjunto não-maximal autoepistêmico. Com isto, $W^* \models \neg O\psi$, pois $w \models \psi$ e $w \notin W^*$. Então, ao considerar W^*, $\Sigma \cup \{\neg O\psi\}$ é satisfatível. Ao considerar apenas os conjuntos maximais autoepistêmicos de atribuição, seria obtido o seguinte resultado: $\Sigma \cup \{\neg O\psi\}$ seria insatisfatível e com isto $\Sigma \models O\psi$. Logo, se qualquer conjunto maximal autoepistêmico satisfaz a Σ, este deve ser equivalente a W e também satisfazer a $O\psi$. Tem-se então que o valor verdade de qualquer fórmula subjetiva dependerá das fórmulas básicas que são acreditadas.

Defina Γ como um **conjunto de crenças** para W se, e somente se, $\Gamma = \{\alpha \mid \alpha$ é básico e $W \models B\alpha\}$. O seguinte teorema determina que fórmulas subjetivas são verdadeiras:

Teorema 4.4.1 (LEVESQUE, 1990): Se W e W^* são conjuntos maximais autoepistêmicos com o mesmo conjunto de crenças, então para qualquer fórmula subjetiva σ, $W \models \sigma$ sse $W^* \models \sigma$.

■

A axiomatização de OL

Resta definir o conjunto de axiomas e regras de inferência de modo a obter as fórmulas válidas de OL. Para o operador O, a intuição existente em $O\alpha$ indica que "α é tudo o que é acreditado", ou seja, "apenas α é acreditado e nada mais". Não é possível então definir $O\alpha$ em função apenas de $B\alpha$ pois, $B\alpha$ indica que "pelo menos α é acreditado". LEVESQUE (1990) então sugere introduzir um

novo operador modal, de modo a indicar que "nada além de α seja acreditado" ("no máximo α é acreditado"). $O\alpha$ será então definido não apenas em função do operador B mas também em função de um novo operador N, de forma que $O\alpha$ será $B\alpha \wedge N\neg\alpha$. Intuitivamente, esta definição pretende indicar que "pelo menos α é acreditado e nada além de α é acreditado". A idéia é ter o operador B como um limite inferior e o operador N como um limite superior do que é acreditado de modo que "tudo o que é realmente acreditado" esteja definido dentro destes limites. Em outras palavras, dado um conjunto de atribuições maximal W e uma fórmula objetiva ϕ, $B\phi$ é verdadeira em relação a W se, e somente se, para todo $w \in W$, $w(\phi) = 1$. Por outro lado, $N\neg\phi$ é verdadeiro quando o conjunto de atribuições satisfazendo ϕ é um subconjunto de W. A semântica para $N\alpha$ é definida do seguinte modo:

$(W,w) \models N\alpha$ sse para todo w', se $(W,w') \not\models \alpha$ então $w' \in W$

LEVESQUE (1990) conclui que o comportamento do operador N é o mesmo do operador de crença B, com a diferença que para N utiliza-se o complemento de W, assim:

$(W,w) \models B\alpha$ sse para todo $w' \in W$, $(W,w') \models \alpha$ e

$(W,w) \models N\alpha$ sse para todo $w' \notin W$, $(W,w') \models \alpha$ (O conjunto de todo $w \notin W$, forma o complemento de W, $\overline{W}$.)

Logo, como $O\alpha$ é definido por $B\alpha \wedge N\neg\alpha$, teremos que, não apenas para $w' \in W$ α é verdadeiro, mas $(W,w) \models N\neg\alpha$ se, e somente se, para todo $w' \notin W$ ($w' \in \overline{W}$), $(W,w') \models \neg\alpha$ ($(W,w') \not\models \alpha$), de forma que se α é verdadeiro em algum mundo w', então $w' \in W$.

Para dar a axiomatização, inicialmente são feitas as

seguintes considerações:

1. No caso de dois agentes, estes serão mutuamente introspectivos
2. Todo o mundo considerado será um elemento de W ou $\overline{W}$

No primeiro caso, estarão axiomas como $N\alpha \supset NN\alpha$ e $N\alpha \supset BN\alpha$. No segundo caso, estará determinada toda fórmula objetiva que é verdadeira em todo mundo membro de $\overline{W}$ e que deve ser falsa em algum membro de W.

Acrescentam-se então aos axiomas anteriores os seguintes axiomas:

1. S5-fraco para B e N:
   a. $B\phi$, onde ϕ é qualquer fórmula objetiva válida
   b. $BC\alpha \supset \beta) \supset (B\alpha \supset B\beta)$, para fórmulas α e β
   c. $\forall x B\alpha \supset B\forall x\alpha$
   d. $(\sigma \supset B\sigma)$, onde σ é uma fórmula subjetiva
   e. os axiomas equivalentes para N;
2. Axiomas que relacionam os operadores N e B:

   $N\phi \supset \neg B\phi$, onde ϕ é uma fórmula objetiva "falsificável"
3. A definição de O:

   $O\alpha \equiv (B\alpha \wedge N\neg\alpha)$, para qualquer fórmula α

Com relação a correção e completude deste conjunto de axiomas,em (LEVESQUE, 1990) tem-se como resultado que este sistema de axiomas é correto e para o caso proposicional ele também é completo. No caso quantificacional, LEVESQUE (1990) acredita que sua axiomatização também é completa, apesar de sua prova usada no caso proposicional não ter funcionado para esse caso.

Algumas aplicações

Serão revistas duas aplicações, entre aquelas originalmente apresentadas. Na primeira será considerado o caso puramente proposicional. Na segunda, serão considerados os quantificadores (LEVESQUE, 1990).

Aplicação 1

Seja α uma fórmula objetiva que é acreditada ser verdadeira e não implica na fórmula ϕ. Nesta primeira aplicação, o que se deseja mostrar é que se a fórmula $O(\alpha \wedge (\neg B\phi \supset \neg\phi))$ é verdadeira, então $B\neg\phi$ também o será.

Com os resultados de "correção e completude", será revista aqui a prova desta propriedade usando os axiomas e regras de inferência:

1. $O(\alpha \wedge (\neg B\phi \supset \neg\phi))$	hipótese
2. $B(\alpha \wedge (\neg B\phi \supset \neg\phi))$	1 e def. de O
3. $(B\neg B\phi \supset B\neg\phi)$	2 e SS-fraco
4. $(\neg B\phi \supset B\neg\phi)$	3 e SS-fraco
5. $N\neg(\alpha \wedge (\neg B\phi \supset \neg\phi))$	1 e def. de O
6. $N(\alpha \supset \phi)$	5 e SS-fraco
7. $\neg B(\alpha \supset \phi)$	6 e o axioma que relaciona N e B
8. $\neg B\phi$	7 e SS-fraco
9. $B\neg\phi$	4, 8 e SS-fraco

■

Aplicação 2

Como uma segunda aplicação, serão considerados os quantificadores. Neste caso, α será uma fórmula (ou uma

conjunção de fórmulas) representando crenças, onde outras crenças poderão ser acrescentadas. Considere então a fórmula abaixo que será apresentada dentro do operador O:

$\gamma = \forall x\ [(\text{pássaro}(x) \wedge \neg B \neg \text{voa}(x)) \supset \text{voa}(x)]$

Suponha α = (pássaro(Tweety)). Prova-se que as seguintes fórmulas são válidas:

1. $O(\alpha \wedge \neg \text{voa}(\text{Tweety}) \wedge \gamma) \supset B \neg \text{voa}(\text{Tweety})$
2. $O(\alpha \wedge \text{voa}(\text{Tweety}) \wedge \gamma) \supset B\text{voa}(\text{Tweety})$
3. $O(\alpha \wedge \gamma) \supset B\text{voa}(\text{Tweety})$

As duas primeiras fórmulas (1 e 2) vem da definição de "tudo que é acreditado" (a noção do "only knowing"). A prova da terceira é feita diretamente utilizando os axiomas e regras de inferência:

1. $O(\alpha \wedge \gamma)$	hipótese
2. $B(\alpha \wedge \gamma)$	1 e definição de O
3. $(\neg B \neg \text{voa}(\text{Tweety}) \supset B\text{voa}(\text{Tweety})$	2 e S5-fraco
4. $N \neg(\alpha \wedge \gamma)$	1 e definição de O
5. $N(\alpha \supset \exists x \neg \text{voa}(x))$	4 e S5-fraco
6. $\neg B(\alpha \supset \exists x \neg \text{voa}(x))$	5 e o axioma N x B
7. $\neg B \neg \text{voa}(\text{tweety})$	6 e S5-fraco
8. $B\text{voa}(\text{Tweety})$	3, 7 e S5-fraco

■

É válido observar a semelhança desta aplicação 2 com o exemplo (4.1.1), onde o fato (2) é representado pela fórmula γ. Mais uma vez, toda a não-monotonicidade se encontra dentro do operador O, o qual traduz o fato (5) do exemplo citado. Com relação as fórmulas (1), (2) e (3) desta segunda aplicação, note também que (3) representa a seguinte fato: se tudo o que acreditamos é que Tweety é um

pássaro e em geral pássaros voam, conclui-se que Tweety voa. Já em (1) e (2) está representado a característica da não-monotonicidade ao acrescentarmos aos fatos conhecidos, o fato de Tweety voar ou não.

Seção 4.5

RELAÇÃO ENTRE OS FORMALISMOS APRESENTADOS

Como já foi dito anteriormente, os formalismos de MOORE (1985) e HALPERN e MOSES (1984b) são bem semelhantes, divergindo apenas nas implicações causadas pelas diferenças entre conhecimento e crença. Comparando esses formalismos tem-se que se uma fórmula é honesta e tem apenas uma expansão estável T, então este conjunto T é único e igual a T^{α}. No entanto, é possível ter fórmulas, tanto honestas como desonestas, que não possuem expansões estáveis bem como fórmulas desonestas possuindo apenas uma expansão estável. Considere o exemplo:

Exemplo 4.5.1 (HALPERN e MOSES, 1984b): Ao considerar o sistema S5-fraco, pela definição de conjunto estável, se um conjunto estável contém a fórmula Bp, este deverá conter p. Mas, a crença em p não nos permite concluir que p seja verdadeiro, ou seja, p não é consequência proposicional de Bp. Considerando então o sistema S5, Cp é uma fórmula honesta, pois o agente conhecendo apenas Cp descreve totalmente seu estado de conhecimento. Por outro lado, $\neg Cp \supset q$ é desonesta enquanto $\neg Bp \supset q$ possui uma única expansão estável. ■

HALPERN e MOSES (1984b), ao considerarem o conhecimento, relacionam a noção de expansão estável à seguinte definição (neste caso, um agente que conhece α deverá também conhecer as fórmulas que são conseqüências em SB de Cα):

Definição 4.5.1: Um conjunto R está **fixado** em uma fórmula α se R é o conjunto de conseqüências proposicionais de:
$\{\beta \mid C\alpha \supset \beta$ é válida em SB, onde β é proposicional$\} \cup \{C\varphi \mid \varphi \in R\} \cup \{\neg C\varphi \mid \varphi \notin R\}$ ■

Com base nesta definição (4.5.1), os seguintes resultados são obtidos (HALPERN e MOSES, 1984b):

1. Dado uma fórmula α, existe um único conjunto R^{α} fixado em α.
2. Se Cα é consistente então R^{α} é estável, caso contrário R^{α} é um conjunto inconsistente contendo todas as fórmulas da linguagem.
3. α é honesta se, e somente se R^{α} é consistente e $\alpha \in R^{\alpha}$.
4. Dado uma fórmula honesta α, $R^{\alpha} = T^{\alpha} = C(K^{\alpha})$.

Logo, é possível relacionar estes dois formalismos, desde que tratemos com fórmulas honestas, onde para uma fórmula honesta α e pelo resultado (4), tenhamos a "igualdade" entre a noção de expansão estável (MOORE, 1985) e o conjunto $C(K^{\alpha})$, onde K^{α} é o modelo correspondente ao estado de conhecimento do agente quando α é "tudo o que é acreditado".

Comparando-se a lógica de LEVESQUE (1990) e a lógica autoepistêmica de MOORE (1985), tem-se os seguintes fatos: LEVESQUE (1990) considera a sua lógica uma generalização quantificacional correta do trabalho de MOORE (1985) e em

segundo lugar, ele afirma que a noção de expansões estáveis pode ser entendida em termos da noção de "tudo o que é acreditado", propriedade esta também já determinada por MOORE (1985).

LEVESQUE (1990), ao contrário de MOORE (1985) (que trabalha com a linguagem proposicional), trabalha com a linguagem de primeira ordem, o que o obriga a usar a consequência de primeira ordem em substituição a consequência tautológica usada por MOORE (1985).

Dada então as características das Teorias Autoepistêmicas Estáveis (seção (4.2)), as quais caracterizam um estado de crença estável, LEVESQUE (1990), no teorema seguinte, relaciona um conjunto estável a um conjunto de crenças.

Teorema 4.5.1 (LEVESQUE, 1990): Seja Γ um conjunto de fórmulas básicas. Γ é estável sse Γ for um conjunto de crenças para algum W. ■

Um outro resultado importante relaciona a noção de "α ser tudo o que é acreditado " à noção de expansão estável definida na seção (4.2):

Teorema 4.5.2 (LEVESQUE, 1990): Para qualquer α básico e qualquer conjunto maximal autoepistêmico de atribuições W, $W \models O\alpha$ sse o conjunto de crenças de W é uma expansão estável de $\{\alpha\}$. ■

Dado o teorema, para que $O\alpha$ seja verdadeiro, é necessário que o que é acreditado seja uma expansão estável

de α, isto é, derivado de α unicamente através da lógica de primeira ordem e da introspecção. Além disso, desde que tenhamos α como uma fórmula honesta, é possível obter um modelo *Kripke* K^{α} (satisfazendo os axiomas de S5), de forma que C(K$^{\alpha}$) corresponderia ao conjunto de crenças de W, onde W $\vdash$ Oα e W é um conjunto maximal autoepistêmico de atribuições.

Seção 4.6

LÓGICA DO CONHECIMENTO "ENVOLVIDO"

A lógica do conhecimento envolvido considera a noção de conhecimento como uma crença verdadeira que é "justificada" e "indestrutível" ou "inviolável".

Desta forma, para que o agente tenha o conhecimento de um fato φ, as seguintes condições devem ser satisfeitas:

1. φ deve ser verdadeiro

2. O agente deve acreditar em φ

3. O agente deve ser capaz de "justificar" o seu conhecimento

4. A crença do agente em φ deve ser "indestrutível" ou "inviolável" no dado instante de tempo.

As condições (1) e (2) são modeladas na abordagem dos mundos possíveis, onde a relação de acessibilidade para o conhecimento deve ser reflexiva (capítulo II). Por "justificar", entende-se associar φ a um conjunto de fórmulas $\alpha_1,\alpha_2,\ldots,\alpha_k$, onde o agente terá o conhecimento de φ se as fórmulas α_1, $\alpha_2,\ldots,\alpha_k$ forem conhecidas . Este conjunto de fórmulas é chamado de **justificativa ou justificativa positiva do agente**. Assuma que este conjunto

de fórmulas é representado por uma única fórmula α tal que $\alpha \equiv \alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_k$. Se este conjunto é vazio, então $\alpha \equiv$ "*true*".

A idéia da crença ser "indestrutível" permite que o agente antecipe todos os fatos que podem invalidar φ e exclua todas as possibilidades da sua crença em φ ser falsa. O instante de tempo é considerado para evitar que, ao acreditar em um fato φ, este fato não mais possa ser falso, ou seja, deseja-se evitar que a crença seja "eterna" o que, intuitivamente , não é uma noção muito aceita. Logo, associado a φ também se encontra um conjunto de fórmulas $\beta_1, \beta_2, \ldots, \beta_m$, de modo que a crença do agente é indestrutível se o agente não tem o conhecimento das fórmulas presentes neste conjunto, em um instante de tempo considerado. Este conjunto é chamado de **justificativa negativa do agente**. Assume-se também que este conjunto é representado por uma única fórmula β tal que $\beta \equiv \beta_1 \vee \beta_2 \vee \ldots \vee \beta_m$. Quando β é vazio, então $\beta \equiv$ "*false*".

Logo, o agente tem o conhecimento envolvido de um fato φ, em um instante de tempo considerado, se ele tem o conhecimento dos fatos presentes no conjunto α e não tem o conhecimento dos fatos presentes em β.

A não-monotonicidade é então capturada já que é possível alterar o conjunto de justificativas (positivas ou negativas) e com isto um fato que é acreditado poderá posteriormente deixar de ser. Esta lógica também evita o problema da omniscIência lógica quando se considera a noção de conhecimento. Será abordado também como este problema é evitado.

Exemplo 4.6.1 (RAO e FOO, 1987a): Suponha que o agente A_1 consulta seu relógio e ele está marcando dez horas. A_1 acredita que são dez horas. Sua crença é justificada com o fato do relógio ser confiável e preciso. Mas, é possível que o relógio, por um motivo qualquer, tenha parado doze horas antes. Apesar da justificativa e o relógio marcar realmente dez horas, A_1 só terá o conhecimento de que são dez horas se o fato do relógio ter parado doze horas antes pertencer ao seu conjunto β. ■

A linguagem

A linguagem é a lógica proposicional modal para m agentes ($m \geq 1$) como definida na seção (2.3), acrescida dos operadores modais P_i de justificativa ou justificativa positiva, N_i de não conhecido ou justificativa negativa e E_i de conhecimento envolvido, $i = 1,\ldots,m$. Logo, se φ, α, β são fórmulas da linguagem, então $P_i(\varphi,\alpha)$, $N_i(\varphi,\beta)$ e $E_i(\varphi,\alpha,\beta)$ também serão, onde $P_i(\varphi,\alpha)$ indica que "α justifica positivamente φ para o agente i", $N_i(\varphi,\beta)$ indica que "β justifica negativamente φ para o agente i" e $E_i(\varphi,\alpha,\beta)$ indica que "o agente i tem o conhecimento envolvido de φ por causa da justificativa positiva α e da justificativa negativa β".

A estrutura

A estrutura é a n-upla $K = (S,\pi,\rho_1,\ldots,\rho_n,j_1,\ldots,j_n)$ onde:

1. S, π, ρ_i, $i = 1,\ldots n$, são definidos na seção (2.4) acrescentando-se que ρ_i é reflexiva, transitiva e euclidiana.

2. j_i, i = 1,...,n, é uma função que atribui a cada agente i, em cada estado s ∈ S e fórmula φ, duas fórmulas correspondentes as justificativas positiva e negativa respectivamente: $j_i(\varphi,s) = \langle\alpha,\beta\rangle$. Por conveniência, duas proposições primitivas são acrescentadas, "*true*" e "*false*", onde para todo t tal que $(s,t) \in \rho_i$, $\pi(true,t) = V$ e $\pi(false,t) = F$, ou seja, o agente conhece implicitamente o "*true*" e é ignorante do "*false*".

Intuitivamente, φ é do conhecimento envolvido para o agente i se α é verdadeiro em todos os mundos possíveis e β é falso em pelo menos um deles.

Sejam φ, ψ, α e β fórmulas da linguagem. Para dar as noções de validade e satisfatibilidade, novamente basta definir para fórmulas α da forma $M\varphi$, onde M é um operador modal (nos outros casos a definição é a usual):

1. $(K,s) \models C_i\varphi$ sse $(K,t) \models \varphi$ para todo t tal que $(s,t) \in \rho_i$

2. $(K,s) \models P_i(\varphi,\alpha)$ sse $j(\varphi,s) = \langle\alpha,\beta\rangle$, para algum β

3. $(K,s) \models N_i(\varphi,\beta)$ sse $j(\varphi,s) = \langle\alpha,\beta\rangle$, para algum α

4. $(K,s) \models E_i(\varphi,\alpha,\beta)$ sse $(K,s) \models C_i\varphi$, $(K,s) \models P_i(\varphi,\alpha)$, $(K,s) \models C_i\alpha$, $(K,s) \models N_i(\varphi,\beta)$ e $(K,s) \models \neg C_i\beta$.

Da propriedade (4) temos que $\models E_i(\varphi,\alpha,\beta) \equiv C_i\varphi \wedge P_i(\varphi,\alpha) \wedge C_i\alpha \wedge N_i(\varphi,\beta) \wedge \neg C_i\beta$, isto é, o agente é logicamente onisciente com respeito ao operador C_i sem o ser com respeito ao operador E_i. As seguintes fórmulas são satisfatíveis, e mostram porque o conhecimento envolvido não sofre o problema da onisciência lógica:

1. $E_i(\varphi,\alpha_1,\beta_1) \wedge E_i(\varphi \supset \psi,\alpha_2,\beta_2) \wedge \neg E_i(\psi,\alpha_3,\beta_3)$ é satisfatível porque as justificativas α_1, α_2, α_3, bem como

β_1, β_2, β_3, são independentes entre si.

2. $E_i(\varphi, \alpha_1, \beta_1) \wedge \neg E_i(\varphi \wedge (\psi \vee \neg\psi), \alpha_2, \beta_2)$ é satisfatível pela mesma razão anterior.

3. $\neg E_i(\varphi \vee \neg\varphi, \alpha, \beta)$ é satisfatível (fórmulas válidas podem não ser de conhecimento envolvido, já que este depende não só do conhecimento de α, mas do não conhecimento de β. Caso β se torne conhecido em algum momento, esta fórmula pode passar a ser verdadeira).

4. $E_i(\varphi, \alpha_1, \beta_1) \wedge \neg E_i(\varphi \wedge (\psi \vee \neg\psi), \alpha_2, \beta_2)$ é satisfatível, mais uma vez pelo fato das justificativas serem independentes entre si.

Axiomatização

Para dar uma axiomatização correta e completa, visto que o operador C_i satisfaz aos axiomas de $S5_m$, acrescenta-se a $S5_m$ os seguintes axiomas:

A.18. $E_i(\varphi, \alpha, \beta) \equiv C_i\varphi \wedge P_i(\varphi, \alpha) \wedge C_i\alpha \wedge N_i(\varphi, \beta) \wedge \neg C_i\beta$, indicando que o agente só tem o conhecimento envolvido de φ se ele tem conhecimento de φ e de sua justificativa positiva e não tem conhecimento de sua justificativa negativa.

A.19. $P_i(\varphi, \alpha) \supset C_i(P_i(\varphi, \alpha))$

A.20. $N_i(\varphi, \beta) \supset C_i(N_i(\varphi, \beta))$

Estes dois últimos axiomas indicam que o agente conhece as suas justificativas positiva e negativa, desde que estas sejam verdadeiras.

O sistema formado por $S5_m$, acrescentado de A.18,, A.19 e A.20 será conhecido como $S5E_m$. Acrescentando-se a seguinte condição a j_i

C.1. Se $ji(\varphi,s) = \langle\alpha,\beta\rangle$ e $(s,t) \in \rho i$ então $ji(\varphi,t) = \langle\alpha,\beta\rangle$

o seguinte teorema é provado:

Teorema 4.6.1 (RAO e FOO, 1987a): O sistema modal SSEm é uma axiomatização correta e completa para mundos reflexivos-simétricos-euclidianos, cuja a função ji satisfaz a condição C.1. ■

prova: Apêndice A. ■

Logo, a noção de conhecimento envolvido de φ depende do conhecimento de uma fórmula α (justificativa) e do não conhecimento de uma fórmula β (justificativa negativa). Sendo assim, é possível capturar a não-monotonicidade desde que a noção de conhecimento de φ seja modificada quando, por exemplo, a fórmula β passe a ser conhecida, ou seja, quando houver uma alteração em uma das justificativas do agente.

Com a noção de tempo, é possível considerar a ocorrência de mudanças no conhecimento do agente. Estas mudanças são causadas por alterações no conjunto de justificativas do agente, o qual estaria diretamente relacionado a variável tempo:

Como o conhecimento do agente é justificado por um conjunto de justificativas, um fato, com o passar do tempo, só deixará de ser conhecido pelo agente, se não for encontrado um conjunto de justificativas alternativo que possa justificar este conhecimento. Existirá então um certo limite na capacidade do agente em permanecer conhecendo uma determinada fórmula, que será maior ou menor dependendo da

capacidade do agente em permanecer conhecendo a fórmula em um dado intervalo de tempo, ou seja, durante o processo de mudança no conjunto de justificativas. Baseando-se neste limite, é possível então dar uma certa ordenação no conhecimento das fórmulas. RAO e FOO (1987a) não garantem porém que esta noção de tempo seja válida em todas as circunstâncias.

Aplicação: Sistema de Manutenção da verdade (TMS - "Truth Maintenance System" (DOYLE, 1979))

Existe a possibilidade de "relacionar" esta lógica ao TMS. O TMS é um programa de apoio ao raciocínio não-monotônico. Ele trabalha na manutenção da consistência entre declarações geradas por outros provadores de teorema, sem gerar novos fatos. Quando uma inconsistência é encontrada, esta é solucionada pelo TMS através da alteração de um conjunto mínimo de crenças.

No TMS, cada declaração ou regra é representada por um nó. Cada nó pode ser transformado em uma fórmula na lógica do conhecimento envolvido.

Considere então um nó do TMS. Este é composto de um **dado** e um **conjunto de justificativas.** O nó do TMS, pode estar em dois estados:

1. IN : Quando se acredita que o nó é verdadeiro; chama-se estes nós de válidos, ou

2. OUT : Quando não se acredita que o nó seja verdadeiro

Um **dado** de um nó é uma fórmula proposicional que representa uma premissa ou uma suposição. O **conjunto de justificativas** formam as razões (outros nós) para se

acreditar em um determinado nó que se encontra no estado IN.

Uma justificativa SL para um nó tem a seguinte forma: (SL <lista IN> <lista OUT>)

A lista IN e a lista OUT são formadas por um conjunto de nós que pode ser vazio. Dado um nó no estado IN, uma justificativa deste nó é válida se, e somente se, cada um dos nós da sua lista IN estiver no estado IN e cada um dos nós de sua lista OUT estiver no estado OUT. As **premissas** são os nós cuja justificativas SL possuem listas IN e OUT vazias, ou seja, não dependem da presença ou da ausência de determinadas crenças. As suposições são os nós onde a lista OUT não é vazia. Essas podem ser interpretadas da seguinte maneira (SILVA, 1990): Os nós da lista IN abrangem as razões que levam o TMS a assumir o nó justificado; os nós da lista OUT representam o critério específico que autoriza tal suposição.

Seja η o conjunto de nós do TMS. Chama-se de **rótulo** do TMS a atribuição de valores IN/OUT a cada um dos nós de η. Seja η_{in} o conjunto de todos os nós do TMS que se encontram no estado IN e η_{out} o conjunto de todos os nós do TMS que se encontram no estado OUT. Diz-se então que um rótulo é válido se cada nó pertencente a η_{in} tem todos os seus nós da lista IN no estado IN e todos os nós da sua lista OUT no estado OUT, em caso contrário o nó pertence a η_{out}. Um nó N e η_{in} é **bem justificado** se ele possui uma lista IN vazia e uma lista OUT vazia ou, todos os seus nós da lista OUT pertencem a η_{out} e todos os seus nós da lista IN são bem justificados. Um rótulo TMS é **bem fundamentado** se todos os

nós no estado IN são bem justificados. Logo, apesar de um conjunto de nós poder ter vários rótulos, o TMS só considera apenas aqueles bem fundamentados.

Existem outras formas de justificar um nó que podem ser encontradas em (DOYLE, 1979). Em (SILVA, 1990) encontra-se uma revisão sobre TMS.

Dado um conjunto de nós do TMS η e considerando o sistema modal $S5E_m$, podemos transformar cada nó de η em uma fórmula da lógica do conhecimento envolvido. Semanticamente, esta transformação irá impor certas restrições na estrutura *Kripke* do sistema modal $S5E_m$. A tabela abaixo (tabela (IV.1)) mostra para um nó N com um dado (fórmula proposicional) φ, para cada uma das possíveis justificativas, a fórmula da lógica do conhecimento envolvido equivalente com a respectiva restrição imposta a estrutura *Kripke*:

Seja $\alpha_1,\ldots,\alpha_k \in \eta_{in}$ o conjunto de nós pertencentes a lista IN e $\beta_1,\ldots\beta_m \in \eta_{out}$ o conjunto de nós pertencentes a lista OUT. Considere $K\eta = (S,\pi,\rho_1,\ldots,\rho_n,j_1,\ldots,j_n)$ a estrutura *Kripke* do sistema modal $S5E_m$.

JUSTIFICATIVAS	FÓRMULA EQUIV.	RESTRIÇÃO SEMÂNTICA
1. $(SL\langle\alpha_1,\ldots,\alpha_k\rangle \langle\beta_1,\ldots,\beta_m\rangle)$	$Ei(\varphi, \alpha_1\wedge\ldots\wedge\alpha_k, \beta_1\vee\ldots\vee\beta_m)$	$Ji(\varphi,s)=(\alpha_1\wedge\ldots\wedge\alpha_k, \beta_1\vee\ldots\vee\beta_m)$ e para todo t tal que $(s,t)\in\rho i,(K\eta,t)\models\varphi$
2. $(SL\ \langle\ \rangle\ \ \langle\ \rangle)$	$Ei(\varphi, true, false)$	$Ji(\varphi,s)=(true, false)$ e para todo t tal que $(s,t)\in\rho i,\ (K\eta,t)\models\varphi$
3. $(SL\ \langle\ \rangle \langle\beta_1\vee\ldots\vee\beta_m\rangle)$	$Ei(\varphi, true, \beta_1\vee\ldots\vee\beta_m)$	$Ji(\varphi,s)=(true,\beta_1\vee\ldots\vee\beta_m)$ e para todo t tal que $(s,t)\in\rho i,\ (K\eta,t)\models\varphi$
4. $(SL\langle\alpha_1\wedge\ldots\wedge\alpha_k\rangle \langle\ \rangle)$	$Ei(\varphi, \alpha_1\wedge\ldots\wedge\alpha_k, false)$	$Ji(\varphi,s)=(\alpha_1\wedge\ldots\wedge\alpha_k, false)$ e para todo t tal que $(s,t)\in\rho i,(K\eta,t)\models\varphi$
5. não tem	$Ei(\varphi, false, true)$	$Ji(\varphi,s)=(false, true)$ e para todo t tal que $(s,t)\in\rho i,\ (K\eta,t)\models\varphi$

Tabela IV.1

Os αi's e os βi's são fórmulas da linguagem visto que, apesar do dado (φ) ser uma fórmula proposicional, a lista IN e a lista OUT contém nós e não dados. O quadro nos diz que atribuir o estado IN para $\alpha_1,\ldots,\alpha_k$ significa que o agente tem o conhecimento implícito de $\alpha_1,\ldots,\alpha_k$. Do mesmo modo, atribuir o estado OUT para $\beta_1,\ldots,\beta_m$ significa que o agente é implicitamente ignorante de $\beta_1,\ldots,\beta_m$. O último caso considerado no quadro indica que, quando não existe a justificativa SL, a fórmula φ nunca será satisfeita. Sendo assim, sua justificativa positiva será o valor *false* e sua justificativa negativa será o valor *true*.

Exemplo 4.6.2 (SILVA, 1990): Considere os seguintes fatos:

1. A - "Tweety é um animal"
2. B - "Tweety é uma ave"
3. C - "Tweety tem asas"
4. nD - "Tweety não voa"

5. abA - "Tweety é um animal anormal em relação a não voar"

6. abB - "Tweety é uma ave anormal em relação a ter asas"

Sejam então as seguintes fórmulas:

N.1. B (por hipótese)

N.2. B ⊃ A

N.3. abB (sem justificativas)

N.4. B ∧ ¬abB ⊃ C

N.5. C ⊃ abA

N.6. A ∧ ¬abA ⊃ nD

Utilizando a tabela (IV.1), a tabela (IV.2) abaixo apresenta as sentenças (N.1 a N.6) como nós do TMS e como fórmulas da lógica de conhecimento envolvido.

SENTENÇA	NÓ DO TMS	FÓRMULA
N.1	B(SL < > < >)	Ei(B, *true*, *false*)
N.2	A(SL<N.1>< >)	Ei(A, Ei(B, *true*, *false*), *false*)
N.3	abB	Ei(abB, *false*, *true*)
N.4	C(SL <N.2> <N.3>)	Ei(C, Ei(A, Ei(B, *true*, *false*), *false*), Ei(abB, *false*, *true*))
N.5	abA(SL<N.4> < >)	Ei(abA, Ei(C, Ei(A, Ei(B, *true*, *false*), *false*), Ei(abB, *false*, *true*)), *false*)
N.6	nD(SL <N.2> <N.5>)	Ei(nD, Ei(A, Ei(B, *true*, *false*)), Ei(abA, Ei(C, Ei(A, Ei(B, *true*, *false*), *false*), Ei(abB, *false*, *true*)), *false*))

TabelaIV.2

O nó N.1 é uma premissa. Os nós N.4 e N.6 são suposições. Partindo da hipótese N.1, que se encontra no estado IN, é possível identificar dois rótulos válidos:

$$R.1 = \langle \eta_{in}, \eta_{out} \rangle = \langle \langle N.1, N.2, N.4, N.5 \rangle, \langle N.3, N.6 \rangle \rangle$$

$$R.2 = \langle \eta_{in}, \eta_{out} \rangle = \langle \langle N.1, N.2, N.3, N.6 \rangle, \langle N.4, N.5 \rangle \rangle$$

Os nós de η_{in} pertencentes a R.1 são bem justificados e consequentemente o rótulo R.1 é bem fundamentado. Já o nó R.2 não é bem fundamentado, já que o nó N.3 pertencente a η_{in} não é bem justificado.

Logo, utilizando a restrição semântica definida na tabela (IV.1), é possível construir uma estrutura *Kripke* $K\eta$ para o conjunto de nós do exemplo (4.6.2). Note que as únicas fórmulas satisfatíveis nesta estrutura serão aquelas pertencentes ao conjunto η_{in} de um rótulo bem fundamentado (RAO e FOO, 1987). Também, como resultado geral, a "correção" e a "completude" do TMS, em relação a lógica do conhecimento envolvido (sistema modal S5E$_m$), estão em (RAO e FOO, 1987a).

Seção 4.7

CONCLUSÕES

Neste capítulo foi capturada a noção da não-monotonicidade visto que, nos quatro formalismos apresentados, sempre que novos fatos são acrescentados ao conjunto de crenças do agente ou ao conjunto de justificativas (para o último formalismo apresentado), novas crenças poderiam surgir.

Com relação aos três primeiros formalismos, estes representam o estado de crença (ou conhecimento) de um agente quando "tudo o que é acreditado (ou conhecido)" é uma fórmula α. Independente da motivação presente, estes três formalismos estão relacionados segundo as diferenças e semelhanças encontradas entre as noções de conhecimento e crença. LEVESQUE (1990) consegue ir um pouco mais além, pois não mais trabalha com a noção de conjuntos de crenças

nem de estados de conhecimento, para tratar com uma lógica de primeira ordem, permitindo quantificadores no contexto modal.

É possível relacionar a lógica do conhecimento envolvido (RAO e FOO, 1987b) à abordagem definida por HALPERN e MOSES (1984b): RAO e FOO (1987b) quando definem a noção de conjunto estável de fórmulas, não apenas modelam o conhecimento implícito do agente, mas acrescentam também a este conjunto o conhecimento explícito do agente. Este conhecimento explícito está relacionado à função j_i (função de justificativa). Neste caso, são acrescentadas às condições de estabilidade já definidas acima, as seguintes condições:

6. Se $j(\alpha,T) = \langle\varphi,\psi\rangle$ então $P(\alpha,\varphi) \in T$ e $N(\alpha,\psi) \in T$.
7. Se $P(\alpha,\varphi) \in T$ então $j(\alpha,T) = \langle\varphi,\psi\rangle$ para algum ψ.
8. Se $N(\alpha,\psi) \in T$ então $j(\alpha,T) = \langle\varphi,\psi\rangle$ para algum φ.
9. $C\alpha \in T$, $P(\alpha,\varphi) \in T$, $C\varphi \in T$, $N(\alpha,\psi) \in T$ e $\neg C\alpha \in T$ sse $E(\alpha,\varphi,\psi) \in T$.

Com estas condições, RAO e FOO (1987b) afirmam que o conhecimento explícito em um conjunto estável de fórmulas não é fechado sob a implicação (é possível ter α e $\alpha \supset \beta$ conhecidas explicitamente em um conjunto estável T, sem que β seja conhecida explicitamente neste conjunto). Também é possível mostrar que nem todas as fórmulas válidas são explicitamente conhecidas em um conjunto estável. Além destas, o fato do conjunto estável ser determinado pelas fórmulas proposicionais que ele contém (HALPERN e MOSES, 1984b) não é mais verdadeiro pois, um agente com um mesmo conjunto de fórmulas proposicionais pode chegar a dois conjuntos estáveis diferentes usando duas funções j_i

diferentes. Logo, tem-se que as fórmulas de um conjunto estável são determinadas não só pelas fórmulas proposicionais que ele contém mas também pela função de justificativa j_i dada a cada fórmula do conjunto estável (RAO e FOO, 1987b).

Salienta-se também que,na abordagem de LEVESQUE (1990), a proporção em que uma fórmula é acrescentada à base de conhecimento, torna-se necessário que as inferências sejam revistas devido a noção do "only knowing", que poderá acarretar algumas mudanças, mesmo que esta fórmula não seja relevante ao contexto. A lógica do conhecimento envolvido (RAO e FOO, 1987b), por sua vez, não possui este problema: é considerado que, se uma fórmula φ pode ser inferida, é porque certas premissas estão presentes (justificativa positiva de φ) e certas exceções estão ausentes (justificativa negativa de φ). Com isto, ao acrescentar ou retirar uma fórmula da base de conhecimento, as inferências só precisarão ser revistas se esta fórmula for adicionada ao conjunto de exceções ou retirada do conjunto de premissas.

CAPÍTULO V

O CONHECIMENTO PARA FORMALIZAR SISTEMAS DISTRIBUÍDOS

Seção 5.1
INTRODUÇÃO

Uma aplicação da lógica do conhecimento se encontra na especificação formal de Sistemas Distribuídos.

Serão apresentados aqui dois exemplos que tratam do problema da comunicação entre os processos do sistema. Inicialmente, através do **problema do "ataque coordenado"**, será mostrado a importância da comunicação entre os processos quando se deseja atingir o conhecimento comum de um fato. Mostra-se então que, quando a comunicação não é garantida, o conhecimento comum não é atingido. No segundo exemplo, será tratada a relação entre conhecimento, ação e comunicação. O problema considerado é o **problema dos "maridos infiéis"**, bem como suas variações. Com o problema original, é analisada a importância da comunicação quando esta é utilizada na transmissão de uma mensagem, a qual define as ações tomadas pelos processos. As variações do problema original, por sua vez, permitem analisar, entre outros, os casos em que a comunicação é síncrona ou assíncrona bem como casos em que o protocolo utilizado pode tolerar falhas. Antes porém de apresentar estes dois exemplos, serão dados alguns conceitos usuais em sistemas distribuídos.

O conceito de Sistemas Distribuídos apareceu devido a necessidade dos usuários em aumentar a velocidade de

processamento dos computadores. A idéia então era caracterizar um tipo de computação chamada de computação paralela. Esta permite várias unidades de processamento, de modo a ter vários elementos processadores separados, os quais trabalham em paralelo, sendo as instruções executadas simultaneamente. Um sistema de processamento distribuído ou **Sistema Distribuído** é um tipo de computação paralela onde a comunicação entre os elementos processadores é feita através da troca de mensagens.

O atual estágio de desenvolvimento deste tipo de computação é devido a dois fatores : o primeiro é o avanço tecnológico muito rápido e redutor de custos, tanto na microeletrônica quanto nas comunicações. O segundo surge das necessidades do usuário, a cada dia exigindo uma maior sofisticação, no que diz respeito a um processamento distribuído. Com isto, os projetistas hoje buscam alcançar características tais como um alto desempenho, no que se refere a modularidade, alta confiabilidade e disponibilidade, além de um compartilhamento efetivo de recursos.

Pode-se então caracterizar um **Sistema Distribuído (SD)** como um conjunto de processos, consistindo de uma arquitetura modular, com um número variável de elementos de processamento, onde a comunicação é feita por troca de mensagens através de uma rede de comunicação. Existe também uma cooperação entre os processos e um gerenciamento do tempo de execução. Para que esta troca de mensagens entre os elementos de um sistema distribuído tenha êxito, é importante que se estabeleça um conjunto de regras e convenções chamadas de **protocolo**.

O projeto de um protocolo distribuído envolve a determinação do comportamento e da interação entre os processos componentes do sistema. Uma das razões da dificuldade em projetar um protocolo distribuído é o fato da troca de mensagens envolver um atraso substancial na sua entrega com relação a velocidade dos processos. Além desta, salienta-se a independência desses processos, cujas ações dependem de informações contidas nos mesmos, e seus comportamentos falhos ou inesperados justificam a complexidade da análise ou projeto de um SD.

A idéia de formalizar o conhecimento, na análise de protocolos distribuídos, surge da possibilidade de considerar os "estados de conhecimento" dos processos em diversos pontos da execução do protocolo. Além disto, existirão situações onde um certo fato é de conhecimento comum entre os processos de uma rede de comunicação. Esta idéia se relaciona, entre outras, com a noção de ações coordenadas onde este conhecimento comum permitirá uma melhor ou pior comunicação entre os processos.

Este capítulo constará de duas partes : a primeira tem por objetivo apresentar um modelo geral para Sistemas Distribuídos e utilizar a lógica do conhecimento como ferramenta para sua especificação formal. Na segunda parte, serão analisados os dois problemas com suas respectivas variações, onde é exposta a importância do conhecimento comum ao tratar com Sistemas Distribuídos. A formalização a ser apresentada é aquela sugerida em (HALPERN e MOSES, 1984a), (MOSES, DOLEV e HALPERN, 1985), (MOSES, 1986) e em (HALPERN, 1985). Na seção (5.2) será descrito um modelo geral de um SD e conterá um modo de atribuir conhecimento

ao SD descrito. Na seção (5.3) será descrito o problema do "ataque coordenado", considerando as implicações causadas quando a comunicação não é garantida ou confiável e no caso da comunicação ser confiável. Na seção (5.4) serão vistas algumas variantes do conhecimento comum, no momento em que se acrescenta a noção de tempo. Na seção (5.5) será descrito o problema dos "maridos infiéis", bem como algumas de suas variações, as quais permitirão analisar o conhecimento comum nos diversos tipos de comunicação a que está submetido o sistema. Por fim, na seção (5.6), serão dadas algumas conclusões. Salientamos que, para ressaltar os resultados, deixamos algumas das demonstrações no apêndice B.

SEÇÃO 5.2

UM MODELO PARA UM SISTEMA DISTRIBUÍDO (SD)

Sistemas Distribuídos e Processos

Um **Sistema Distribuído** é definido como um conjunto de m processos, $m \geq 2$, ligados por uma rede de comunicação. A comunicação entre os processos se dá através do envio de mensagens entre as linhas da rede. Os processos são os estados da máquina, podem ou não ter *clocks*, que são relógios, onde um relógio é uma função monotônica, não decrescente e de tempo real. Cada processo possui sua história, que é representada por uma sequência finita de mensagens enviadas ou recebidas até um dado tempo real. Esta história é chamada de **história da mensagem** de um processo. Caso o processo possua *clock*, a sequência também é marcada pelo tempo nesse *clock*. Tanto o tempo do *clock* como a história da mensagem de um processo são determinados

em função do tempo real, através da **Função do tempo de clock** e da **função da história da mensagem**, respectivamente. Então, a função do tempo de *clock* indicará quando cada mensagem foi enviada ou recebida e a função da história da mensagem indicará as mensagens enviadas e recebidas pelo processo.

Os processos podem se encontrar em dois estados distintos: **"ativo "** ou **"não-ativo"**. No estado *não-ativo* , o processo se encontra desligado da rede, ou seja, não há trocas de mensagens até que ele passe para o estado ativo. O momento em que o processo passa para o estado ativo é chamado de **tempo inicial do processo** (tempo real) e, neste caso, o processo se encontra em seu **estado inicial**, onde a sua história da mensagem é vazia e, é deste estado que se começa a seguir um determinado protocolo.

Protocolos

Um **protocolo** é uma função que especifica as ações que devem ser tomadas por um processo, em um determinado "ponto", como função de seu estado inicial, de sua história da mensagem e do intervalo de valores lido por seu *clock* (caso este possua) desde o momento em que este passou para o estado ativo. Podem-se citar dois tipos de protocolos: **protocolos determinísticos** e **protocolos não-determinísticos**. Neste caso serão considerados apenas os protoclos determinísticos. Um protocolo não-determinístico é considerado como um conjunto de protocolos determinísticos, cada um correspondendo a uma sequência particular de escolhas não determinísticas (MOSES, 1986). Se G é um conjunto de processos, é possível

ter um protocolo associado a este conjunto G, um *joint* **protocolo**, que seria uma n-upla consistindo de n protocolos, um para cada processo de G.

Runs

Define-se uma *run* r em um SD como sendo a descrição completa do comportamento do sistema. Nesta descrição, estão incluidas a função da história da mensagem, a função do tempo de *clock*, o tempo em que cada processo teve início, o protocolo seguido por cada processo, bem como seu estado inicial. Desta forma, em um determinado tempo t e em uma determinada *run* r, cada processo terá a sua história contendo o seu estado inicial e uma sequência de eventos observados nesta *run* até o tempo t,onde estes eventos são as mensagens enviadas e recebidas. O par (r,t) é definido como um **ponto**, onde r é uma *run* e t é um número real correspondente ao tempo real. Serão consideradas apenas as *runs* consistentes, ou seja, aquelas em que qualquer ação tomada por cada processo está especificada no seu protocolo.

Pode-se então identificar um SD como um conjunto de possíveis *runs*. Este conjunto pertenceria a um protocolo particular, implementado em um determinado ambiente, teria certas propriedades, e conteria todas as informações relevantes ao sistema. O comportamento relativo dos *clocks* e as propriedades do processo de comunicação, entre outras, refletiriam diretamente nas propriedades deste conjunto. Um protocolo, por sua vez, será executado neste conjunto caso exista uma *run* que permita a sua execução.

Estado local e estado global

Dado uma *run* r, é possível atribuir um estado l_i^r a cada processo, chamado de **estado local**. O estado local é função da história da mensagem e consta de todas as mensagens enviadas e recebidas bem como as transições internas do processo até um determinado tempo. Desta forma, também é definida a **visão** de um processo em um dado ponto como sendo tudo aquilo que foi observado pelo processo naquele ponto na *run*. Mais precisamente, define-se v(pj,r,t), a **visão** do processo pj no ponto (r,t) como:

i. Uma **visão inativa** se t ≤ to(pj,r), onde to(pj,r) é o tempo de início do processo pj na *run* r, ou

ii. Como um conjunto consistindo do estado inicial do processo pj, da história da mensagem e do intervalo de valores lidos por seu *clock* desde o estado ativo, junto ao protocolo seguido em r.

Pode-se então capturar a idéia de um **estado global**: Dado um ponto (r,t), um **estado global** G(r,t) de um SD é uma n-upla $\langle l_1^r,\ldots,l_n^r\rangle$ de estados locais de cada processo naquele ponto, onde n é o número de processos ativos no ponto (r,t). Pode-se acrescentar a esta n-upla mais um elemento que representa o estado do ambiente, ou seja, o estado contendo as características importantes do sistema não presentes no estado local dos processos. Este foi excluido já que não influi nos resultados que serão apresentados. Se em dois pontos (r,t) e (r',t') o processo pi possui um mesmo estado local em G(r,t) e G(r',t') então estes pontos são **não-distintos** e denota-se por (r,t) ≈i (r',t'). Logo, se dois pontos são não-distintos para um

processo pi, as ações executadas por pi nestes pontos serão as mesmas, já que estas executam um protocolo em função do seu estado local.

Atribuição de conhecimento aos processos

Resta agora definir como atribuir conhecimento aos processos em um SD, ou seja, definir o que significa para um processo conhecer um fato φ. Existem várias maneiras de atribuir conhecimento aos processos em um SD, que variam de acordo com a aplicação a qual o sistema está submetido, isto é, dependendo da aplicação, é possível encontrar uma interpretação de conhecimento apropriada. Neste caso, o que se deseja é definir uma interpretação onde, intuitivamente, um processo conhece um fato φ se seu estado implica na veracidade de φ. Além disso, independente da interpretação dada, a noção de conhecimento em um SD deverá satisfazer as seguintes propriedades:

1. O conhecimento do processo em um dado ponto deve ser função da visão do processo naquele ponto, isto é, quando um processo possui uma mesma visão em dois pontos distintos, este terá o conhecimento dos mesmos fatos nos dois pontos.

2. Apenas os fatos verdadeiros são conhecidos.

As noções de interpretação aqui definidas estão em (HALPERN e MOSES, 1985). Segundo HALPERN e MOSES (1985), o fato de estarmos tratando com sistemas distribuídos nos leva a definir o conhecimento a partir do sistema considerado. Por esta razão, são definidas as diversas interpretações, de maneira bem geral, de modo a permitir que, em qualquer aplicação particular, uma interpretação

particular seja escolhida. Intuitivamente, parece coerente definir o conhecimento partindo do modelo do sistema no lugar de começar de algum modelo teórico de conhecimento e "adequar" este modelo ao sistema.

A linguagem

Antes porém de definir as diversas noções de interpretação, será introduzida a linguagem. A linguagem considerada é uma linguagem proposicional de conhecimento. Os operadores de conhecimento (conhecimento comum, conhecimento implícito, etc.) são usados quando necessários. Os fatos básicos do sistema são representados pelos símbolos proposicionais.

Interpretação Epistêmica

Uma **Interpretação Epistêmica** para o sistema SD é, intuitivamente, a especificação de tudo o que o processo tem conhecimento em um dado ponto, como função da visão do processo relativa a sua história naquele ponto. Esta interpretação vai considerar que a noção de conhecimento obedeça ao sistema modal S5.

Formalmente, uma **Interpretação Epistêmica I** é uma função atribuindo a todo processo pj, em qualquer ponto (r,t), um conjunto de fórmulas C_j^I (r,t). Estas fórmulas pretendem ser os fatos que pj tem conhecimento. Logo, C_j^I(r,t) precisa ser função da visão de pj em (r,t), isto é, se v(pj,r,t) = v(pj,r',t'), para dois pontos (r,t) e (r',t'), então C_j^I (r,t) = C_j^I (r',t').

Dada uma fórmula φ da linguagem, uma Interpretação de conhecimento I e um ponto (r,t), diz-se que φ é verdadeira

em (r,t), sob esta interpretação, (I,r,t) $\models \varphi$, de acordo com os seguintes casos:

1. φ é uma proposição primitiva : Neste caso, (I,r,t) $\models \varphi$ sse (r,t) $\in \rho(\varphi)$, onde $\rho(\varphi) \subseteq S \times (-\infty,\infty)$.*

2. φ é uma conjunção ou negação: (I,r,t) $\models \varphi$ é definido como no caso proposicional.

3. φ é da forma Ci(α): (I,r,t) $\models \varphi$ sse $\alpha \in C_i^I(r,t)$

4. φ é da forma CCg(α): Neste caso, (I,r,t) $\models$ CCg(α) sse para todas as sequências finitas $pi_1, pi_2, \ldots, pi_n$ de processos em g, (I,r,t) $\models$ (Ci1)(Ci2)...(Cin)(α).

Para garantir que uma interpretação epistêmica satisfaça o sistema modal S5 tem-se que, para todos os processos pi, tempo t, runs r $\in$ S e fórmulas da linguagem, se (I,r,t) $\models$ Ciφ então (I,r,t) $\models \varphi$, ou seja, o axioma A.3 (Ci$\varphi \supset \varphi$) é satisfeito para as *runs* de S. Tais interpretações epistêmicas são chamadas de **Interpretações de Conhecimento**

Define-se que um processo pi **suporta** CCg(φ) em (I,r,t) se pi tem o conhecimento de todas as fórmulas que constituem CCg(φ), ou seja, (I,r,t) $\models$ Ci((Ci1)(Ci2)...(Cin)(φ)) (para todas as sequências $pi_1, \ldots, pi_n$ de processos em G). É fácil ver que o seguinte lema vale para interpretações de conhecimento:

lema 5.2.1 (Moses,1986): Se I é uma interpretação de conhecimento para S e r $\in$ S, então, qualquer que seja pi $\in$ g, para todo ponto (r,t), pi suporta CCg(φ) em (I,r,t) sse (I,r,t) $\models$ CCg(φ).

prova: Apêndice B ■

---

* Cada proposição p está associada a um conjunto de pontos $\rho(p)$ que representa os pontos em que p é verdadeiro.

Interpretações baseadas em estado

Uma forma de atribuir conhecimento aos processos é através das **interpretações de conhecimento baseadas em estado.** Neste tipo de interpretação, os processos são analisados considerando o estado e o tempo em que eles se encontram. Numa interpretação de conhecimento baseada em estado I_e, um estado $s(p_j,r,t)$ está associado a um dado ponto (r,t) e é função da visão de p_j em (r,t). Para estender esta noção de estado a um grupo g de processos, basta tomar o estado de g em um dado ponto como sendo os estados dos membros de g naquele ponto. Formalmente: $s(g,r,t) = \{\langle p_i, s(p_i,r,t)\rangle \mid p_i \in g\}$.

Formalmente, uma interpretação de conhecimento I_e é uma **interpretação baseada em estado** se ela satisfaz a: $\varphi \in C_j^{I_e}(r,t)$ sse $(I_s,r',t') \models \varphi$ para todos os pontos (r',t') com $r' \in S$, tais que $s(p_j,r,t) = s(p_j,r',t')$. Pode-se então concluir que:

$(I_s,r,t) \models C_i\varphi$ sse $(I_s,r',t') \models \varphi$ para todo (r',t') tal que $s(p_i,r,t) = s(p_i,r',t')$.

$(I_s,r,t) \models CT_g(\varphi)$ sse para todo $p_i \in g$ e para todo (r',t') tal que $s(p_i,r,t) = s(p_i,r',t')$, $(I_s,r',t') \models \varphi$.

Para dar a noção de conhecimento comum, define-se a relação indicando quando um ponto (r',t') é **acessível** de um ponto (r,t) (seção (2.7)). Isto acontece se existem pontos $(r_1,t_1),\ldots,(r_m,t_m)$ com $r_i \in S$ e processos $p_{j1},\ldots,p_{jm-1} \in g$, tais que $(r,t) = (r_1,t_1)$, $(r_m,t_m) = (r',t')$ e $s(p_{ji},r_i,t_i) = s(p_{ji},r_{i+1},t_{i+1})$ para $1 \leq i \leq m$. Logo, para o conhecimento comum tem-se que:

$(I_s,r,t) \models CC_g(\varphi)$ sse $(I_s,r',t') \models \varphi$ para todo (r',t') acessível de (r,t).

Vale observar que, dado uma Interpretação de conhecimento baseada em estado e um SD, esta interpretação para o SD satisfaz os axiomas do sistema S5. Além disso, para atribuir conhecimento ao processo, este não precisa necessariamente ter "consciência" do seu conhecimento e nem executar qualquer ação particular para chegar a este conhecimento. Em particular, se existe um único estado s(pj,r,t), para todo pj, todas as *runs* r e todo t, então todo fato verdadeiro em todas as *runs*, para qualquer que seja t, é de conhecimento comum a todos os processos sob esta interpretação e, neste caso, as noções de conhecimento comum, conhecimento implícito e conhecimento de todos são equivalentes, o que diverge da hierarquia dada na seção (2.7).

Interpretação de visão total

É possível definir um outro tipo de interpretação, a **Interpretação de visão total**, onde o estado $s(p_j,r,t)$ é definido como sendo igual a v(pj,r,t). Nesta interpretação, os processos não esquecem os fatos conhecidos enquanto que na interpretação baseada em estado, se é possível para um processo chegar a um estado através de duas histórias da mensagem diferentes então, neste estado, não é possível fazer uma distinção entre estas duas histórias "passadas". Isto permite distinguir as visões da melhor maneira possível. Anteriormente, como o estado s(pj,r,t) era função da visão de pj em (r,t), então, se v(pj,r,t) = v(pj,r',t') então, necessariamente teríamos que s(pj,r,t) = s(pj,r',t').

Sistemas Distribuídos e Estruturas Kripke

Para relacionar um SD a uma estrutura *Kripke*, basta tomar os pontos (r,t) do sistema como sendo os mundos possíveis e a relação $\approx_i$ (não-distinguibilidade) como a relação de acessibilidade (note que $\approx_i$ é uma relação de equivalência). Os fatos básicos do sistema seriam as proposições primitivas. Para dar o valor verdade de uma fórmula em um ponto (r,t) de um sistema SD, cada sistema SD é associado a uma estrutura *Kripke* $K_{SD} = (S, \pi, \rho_1, \ldots, \rho_n)$, onde:

- S é o conjunto de pontos (r,t) de SD
- π é uma função que associa a cada ponto (r,t) o valor verdade de uma proposição primitiva p, onde, $\pi(r,t,p) = V$ sse $(r,t) \in \rho(p)$.
- ρ_i é a relação de acessibilidade, onde, para cada processo i, $((r,t),(r',t')) \in \rho_i$ sse $(r,t) \approx_i (r',t')$, ou seja, $l_i^r = l_i^{r'}$.

Dado uma fórmula φ da linguagem, φ é satisfeita em um ponto (r,t) de um sistema SD, sob uma interpretação de conhecimento baseada em estado $I,(I,r,t) \models \varphi$ se, e somente se, ela for satisfeita no estado correspondente da estrutura *Kripke* K_{SD}.

Nas próximas seções serão analisados os dois exemplos. Na realidade, será feita uma análise do conhecimento comum e sua relação com as propriedades inerentes ao processo de comunicação.

Seção 5.3

O PROBLEMA DO ATAQUE COORDENADO

O primeiro exemplo a ser analisado é o problema do **"Ataque coordenado"**. Neste problema, é possível tratar a influência da comunicação para se atingir o conhecimento comum de determinados fatos, em termos de confiabilidade, além disso, é possível analisar os resultados obtidos quando se considera a variável tempo.

O interesse por este problema no estudo de sistemas distribuídos está no fato de, através dele, ser possível capturar a importância da comunicação entre os elementos do grupo no momento em que se deseja atingir o conhecimento comum de um fato. O problema é descrito em [MOSES, 1986] da seguinte forma:

"Suponha que duas divisões de um exército estão posicionadas respectivamente em dois topos de montanha a observar um inimigo que se encontra no vale. Sabe-se que, para vencer a batalha, as duas divisões devem atacar o inimigo simultaneamente. Como não havia planos para o ataque, o general da 1ª divisão idealizou coordenar um ataque simultâneo , em algum momento do dia seguinte. Neste caso, nenhum dos dois generais atacaria o inimigo sem a certeza de que o outro general também estaria atacando . A comunicação entre os generais é feita através de mensageiros que normalmente, levam uma hora para sair de uma divisão e chegar a outra. Com isto, é fácil aceitar que existe a possibilidade do mensageiro se perder na escuridão ou, até mesmo, ser capturado pelo inimigo". A pergunta é: Quanto tempo será necessário para coordenar este ataque?

Para responder esta pergunta, suponha que o general da

1ª divisão, general A, envie ao general B, da 2ª divisão a seguinte mensagem: "Atacar a meia noite". O general B não irá atacar, visto que A não tem o conhecimento de que B recebeu a mensagem. Desta forma, B envia um mensageiro que confirme que a mensagem foi recebida. Mais uma vez, A não irá atacar sem que seja confirmado a B que o seu mensageiro chegou com a mensagem. Seguindo este raciocínio, observa-se que sempre haverá a necessidade de uma confirmação. Logo, os generais nunca irão chegar a fazer um ataque simultâneo.

A prova desta afirmação é feita por indução no número n de mensagens enviadas. Para n = 0, nenhuma mensagem foi enviada e com isto, B não sabe do ataque, sendo impossível o ataque simultâneo. Assuma então que n mensagens não são suficientes, deseja-se provar que para n+1 mensagens também não é possível o ataque simultâneo. Se n+1 mensagens forem suficientes, significa que aquele general que enviou a mensagem de número n+1 irá atacar sem ter o conhecimento de que esta foi realmente entregue, já que não recebeu a sua confirmação. Logo, esta n+1-ésima mensagem não seria necessária e o ataque poderia ser feito com n mensagens, o que é uma contradição, por hipótese.

Em termos de conhecimento, ao receber uma mensagem, o general aumenta em um nível o seu conhecimento sobre o ataque desejado. Isto indica que quando o general B recebe a mensagem m, tem-se $C_B m$. Quando A recebe a confirmação de B, tem-se $C_A C_B m$. Na próxima confirmação tem-se $C_B C_A C_B m$, e, desta forma, os generais não conseguem chegar a um acordo quanto ao ataque coordenado com um número finito de mensagens.

Sistemas onde a comunicação não é garantida ou confiável

Na realidade, a dificuldade em coordenar um ataque se encontra na impossibilidade de garantir a entrega da mensagem, ou seja, no fato da comunicação entre os generais não ser garantida. Intuitivamente, pode-se dizer que se a comunicação não é garantida ou confiável, então possivelmente em algum momento, alguma mensagem não será entregue. Para formalizar este fato, considere a definição (MOSES, 1986):

Definição 5.3.1: (Uma *run* estendida)

Uma *run* r' **ESTENDE** a um ponto (r,t) se em r' e em r, todos os processos possuirem os mesmos estados iniciais, o mesmo tempo inicial, além de suas funções de tempo de *clock* e da história da mensagem serem idênticas até um tempo t (sem incluí-lo). Logo, se (r,t) e (r',t) possuem histórias idênticas, r' estende (r,t). ■

Dado então um sistema S e um conjunto M de mensagens (não vazio), para que a comunicação em S não seja confiável, é preciso que: para toda *run* r ∈ S, tempo t e processo p_i, exista uma *run* r' ∈ S que estenda (r,t) de modo que em (r',t) o processo p_i não receba nenhuma mensagem pertencente a M (mas receba todas as mensagens m' ∉ M) que ele recebe em (r,t). Além disso, todo processo p_j (diferente de pi) recebe as mesmas mensagens em (r,t) e (r',t) e nenhuma mensagem é entregue em r' após o tempo t.

A comunicação não será garantida em um sistema S se, (r,t') e (r',t') forem idênticos para todo t' < t, e, em t, todos os processos pj (diferentes de pi) recebem as mesmas

mensagens em r e r', com exceção feita a pi que em r' deixa de receber aquelas mensagens pertencentes ao conjunto M, além do que nenhuma outra mensagem é entregue em r' após o tempo t. Desta forma, será sempre possível que nenhuma mensagem seja entregue a partir de um certo tempo.

Formalmente, a impossibilidade dos generais coordenarem o ataque é justificada relacionando este problema com o problema de atingir o conhecimento comum.

Define-se uma fórmula φ como uma **fórmula indeterminada**, em um sistema S, no ponto (r,t), se, para alguma *run* r' $\in$ S que estende (r,t), φ não é verdadeira em (r',t') para todo t' $\geq$ t (em caso contrário, a fórmula é determinada). Considere então o seguinte lema:

Lema 5.3.1 (MOSES, 1986): Seja S um sistema onde a comunicação não é garantida e seja I uma interpretação de conhecimento para S. Sejam r,r_1 $\in$ S *runs* tais que r estende (r_1,t_1) e seja t $\geq$ t_1. Então para todas as fórmulas φ, (I,r_1,t) $\models$ $CC_G(\varphi)$ sse (I,r,t) $\models$ $CC_G(\varphi)$. ■

prova : Apêndice B. ■

Teorema 5.3.1 (MOSES, 1986): Seja S um sistema onde a comunicação não é garantida tal que $CC_G(\varphi)$ é indeterminada em S em (r_1,t_1). Se r $\in$ S estende (r_1,t_1) e t $\geq$ t_1 então, em (r,t), $CC_G(\varphi)$ não é verdadeira. ■

prova: Apêndice B. ■

A seguinte proposição justifica a impossibilidade dos generais coordenarem o ataque:

proposição 5.3.1:(MOSES, 1986): No problema do ataque coordenado, qualquer protocolo que garanta que se uma divisão atacar, então as duas divisões atacarão

simultaneamente, é um protocolo em que necessariamente nenhuma das divisões irá atacar. ■

prova: Apêndice B ■

Sistemas onde a comunicação é garantida

Além do resultado obtido na proposição (5.3.1), é possível também mostrar que em sistemas onde a comunicação é garantida, o conhecimento comum pode não ser atingido desde que exista alguma incerteza no tempo necessário para a mensagem ser entregue. Intuitivamente, o seguinte exemplo dá esta noção:

Exemplo 5.3.1: Considere um sistema composto por dois processos p_1 e p_2 conectados por uma linha de comunicação. Estes processos usam *clocks* comuns e a comunicação é garantida, de modo que, qualquer mensagem enviada de p_1 para p_2 atinge p_2 no máximo após ε segundos. Seja f um fato indicando que uma mensagem m foi enviada. Para analisar as mudanças do estado de conhecimento de f em função do tempo, considere um tempo t_2 em que a mensagem m foi entregue a p_2 e um tempo t_1 em que p_1 enviou a mensagem para p_2. É fácil ver que em t_2 tem-se a validade de C_2f, ou seja, p_2 conhece f, e como no máximo em ε unidades de tempo m será entregue, em $t_1+\varepsilon$ tem-se também a validade de C_2f. Também, p_2 não terá o conhecimento de que p_1 tem o conhecimento de que p_2 conhece f antes de $t_2 + \varepsilon$. Logo, p_1 espera $t_1 + 2\varepsilon$ para ter o conhecimento de que $t_2 + \varepsilon$ passou. Seguindo este raciocínio, pode-se chegar a validade da fórmula $(C_1C_2)^n f$ em $t_1 + n\varepsilon$. Como $(CC_g)f$ implica em $(C_1C_2)^n f$ para $n = 1,2,3,\ldots$, $(CC_g)f$ nunca será atingido.

■

Acrescentando, ao fato f, uma informação relacionada ao tempo em que a mensagem m foi enviada, são obtidos os seguintes resultados: suponha que f é um fato indicando que a mensagem foi enviada em um tempo t_1. Neste caso, no instante $t_1 + \varepsilon$, tanto $C_1 f$ quanto $C_2 f$ são válidos e o conhecimento comum neste caso pode ser atingido. Isto significa que houve uma modificação simultânea da visão dos processos.

Seção 5.4

VARIAÇÕES DO CONHECIMENTO COMUM

No problema do ataque coordenado, é possível atingir estados de conhecimento, os quais são variações do conhecimento comum. Em certo sentido, estes estados são considerados mais "fracos" que o conhecimento comum por serem mais facilmente atingidos. Será dado então uma nova visão do conhecimento comum.

É possível considerar uma visão diferente de conhecimento comum sob uma interpretação de conhecimento baseada em estado. Por exemplo, suponha o anúncio de um fato φ a um grupo g de agentes. Suponha também que é de conhecimento comum que todos em g, após o anúncio, compreendem φ simultaneamente. Logo, $(C^{C}_{g})\varphi$ é atingido. No entanto, ao atingir o conhecimento comum de φ, os agentes de g não precisariam conhecer separadamente a sequência infinita da forma $(C^{T}_{g})^{n}\varphi$ que aparece na definição de $(C^{C}_{g})\varphi$. Isto pode significar que, após o anúncio, os membros de g passaram para um estado de conhecimento X (onde $(C^{C}_{g})\varphi$ foi atingido) caracterizado pelo fato de φ ser verdadeiro e

todos em g terem o conhecimento de que X é verdadeiro. A seguinte equação é satisfeita:

$$X \equiv \varphi \wedge (C_{T_G})X$$

Ao considerar uma interpretação de conhecimento baseada em estado, I_e, o seguinte teorema é válido:

teorema 5.4.1 (MOSES, 1986): Seja I_e uma interpretação baseada em estado para um sitema S. Para todas as *runs* r ∈ S e todo tempo t,

$$(I_e, r, t) \models (C_{C_G})\varphi \equiv \varphi \wedge (C_{T_G})(C_{C_G})\varphi$$

■

Logo, neste tipo de interpretação, $(C_{C_G})\varphi$ é a solução "mais fraca" para a equação $X \equiv \varphi \wedge (C_{T_G})X$ no sentido de que para qualquer solução dada a essa equação, $X \supset (C_{C_G})\varphi$ é verdadeira (Note que $C_{C_G}(\varphi \wedge \psi)$ é solução para esta equação). Assim, para interpretações de conhecimento baseadas em estado em que φ é verdade, $C_{C_G}(\varphi)$ é um "ponto fixo" do operador C_{T_G}*. Também, nessas interpretações, os seguintes axiomas são válidos para o conhecimento comum:

1. Axioma do ponto fixo : $(C_{C_G})\varphi \supset \varphi \wedge (C_{T_G})(C_{C_G})\varphi$
2. Axioma da indução : $(C_{C_G})(\varphi \supset (C_{T_G})\varphi) \supset (\varphi \supset (C_{C_G})\varphi)$
3. Axioma do fecho : $((C_{C_G})\varphi \wedge (C_{C_G})(\varphi \supset \psi)) \supset (C_{C_G})\psi$

ε-conhecimento comum $(C_{C_G})^{\varepsilon}$ e conhecimento comum eventual $(C_{C_G})^{\hat{\forall}}$

Como atingir o conhecimento comum é problemático em SD's (proposição (5.3.1)), pode-se indagar que outro tipo de conhecimento pode ser atingido ao considerar SD's onde há garantia da comunicação, mas existem incertezas no tempo

---

* Detalhe sobre a noção de ponto fixo podem ser encontrados em (MOSES, 1986).

de transmissão da mensagem. Para tal, estende-se a linguagem de modo a permitir os operadores temporais $\hat{\Diamond}\varphi$ significando "eventualmente φ" e $O^{\delta}\varphi$ significando "Em δ unidades de tempo a partir de agora, φ será verdadeira". Além destes, pode-se definir $\Box\varphi$ significando "sempre φ" onde $\Box\varphi$ seria definido como $\neg\hat{\Diamond}\neg\varphi$. Neste caso adiciona-se mais uma cláusula ao considerar a validade de uma fórmula da linguagem (seção (5.2)): se φ é uma fórmula e δ é um número real, $\hat{\Diamond}\varphi$ e $O^{\delta}\varphi$ são fórmulas. Na definição semântica, são acrescentadas:

1. $(I,r,t) \models \hat{\Diamond}\varphi$ sse para algum $t' \geq t$, $(I,r,t') \models \varphi$, e
2. $(I,r,t) \models O^{\delta}\varphi$ sse $(I,r,t+\delta) \models \varphi$

Visto que estamos tratando com SD em que a comunicação não é instantânea, os fatos serão considerados sempre estáveis, ou seja, uma vez verdadeiros, permanecerão verdadeiros para sempre. Logo, se φ é estável, $\varphi \supset \Box\varphi$ é válido. Da mesma forma, se φ é estável, $\hat{\Diamond}\varphi$ e $O^{\delta}\varphi$ também o serão. Mas nem sempre o conhecimento de fatos estáveis é estável. Para observar isto, basta considerar uma interpretação de conhecimento baseada em estados, I_s, um fato estável φ e um processo p_i. É possível que em um tempo t, $C_i\varphi$ seja verdadeiro em I_s, mas em $t' \geq t$ isto não aconteça devido ao novo estado em que p_i se encontra. Casos como esse não ocorrem nas interpretações de visão total, onde um fato é estável se, e somente se, todos os processos têm o conhecimento de que ele é estável e não esquecem os fatos que eles têm conhecimento. Logo, nesta última interpretação, se φ é estável, $C_i\varphi$, $(CT_G)\varphi$ e $(CC_G)\varphi$, etc. também o serão. Por isso, no que se segue, serão consideradas apenas as interpretações de conhecimento de

visão total.

Difusão síncrona

Considere um sistema onde toda mensagem enviada é entregue a todos os processos $p_i \in S$ em um intervalo de tempo de ε unidades, a partir do tempo real em que a mensagem foi enviada. Como se vai trabalhar com interpretação de visão total, considere também que todas as propriedades do sistema são conhecidas dos processos. Seja então o estado de conhecimento do sistema quando um processo p_i recebe uma mensagem m: ao receber m, p_i adquire o conhecimento de um fato f = **"envio de m"** e também tem o conhecimento de que dentro de ε unidades de tempo todos terão o conhecimento de f e que, com mais ε unidades de tempo, todos os processos terão o conhecimento de que todos os processos têm o conhecimento de f e assim por diante. Desta forma, $(CC_G)^{\varepsilon}\varphi$ é definido como o maior ponto fixo da equação:

$$(CC_G)^{\varepsilon}\varphi \equiv \varphi \wedge O^{\varepsilon}(CT_G)(CC_G)^{\varepsilon}\varphi$$

Logo, f se torna de ε-conhecimento comum assim que m seja entregue a qualquer processo. Comparando-se as noções de conhecimento comum e ε-conhecimento comum, pode-se constatar que o conhecimento comum é um estado de conhecimento estático, independente do tempo, enquanto que o ε-conhecimento comum é uma noção que depende essencialmente de tempo (isto é, $(CC_G)\varphi$ pode ser verdade num tempo t independente de seu valor verdade em $t' < t$ ou em $t' > t$ enquanto que, se $(CC_G)^{\varepsilon}\varphi$ é verdade ou não, irá depender do que os processadores vão saber dentro de ε

unidades de tempo, 2ε, etc).

Difusão assíncrona

Considere um sistema de difusão usando canais de comunicação assíncrona. Neste sistema, toda mensagem enviada irá **eventualmente** atingir todos os processos. Como exemplo, seja f = "envio de m", desta forma, ao receber m, o processo terá o conhecimento de f e terá o conhecimento de que eventualmente todos os outros processos irão receber m e terão o conhecimento de f e que eventualmente todos os outros processos irão receber m e assim por diante.

Este estado de conhecimento bem mais fraco que os dois anteriores é chamado de conhecimento comum eventual ou **$\hat{\diamond}$-conhecimento comum**, $(CC_g)^{\hat{\diamond}}$, onde o operador $\hat{\diamond}$ corresponde ao "eventualmente". Logo, $(CC_g)^{\hat{\diamond}}$ é definido como o maior ponto fixo de :

$$(CC_g)^{\hat{\diamond}}\varphi \equiv \varphi \wedge (CT_g)^{\hat{\diamond}}(CC_g)^{\hat{\diamond}}\varphi \ ,$$

onde $(CT_g)^{\hat{\diamond}} \equiv \bigwedge_i C_i^{\hat{\diamond}}\varphi$ e $C_i^{\hat{\diamond}}\varphi$ indica que eventualmente o processo p_i conhece o fato φ.

Os axiomas anteriormente citados são válidos tanto para o ε-conhecimento comum como para o conhecimento comum eventual. Basta apenas trocar, no primeiro caso, CT_g por $O^{\varepsilon}(CT_g)$ e CC_g por $(CC_g)^{\varepsilon}$ e no segundo caso, CT_g por $(CT_g)^{\hat{\diamond}}$ e CC_g por $(CC_g)^{\hat{\diamond}}$. Pode-se também observar que como no caso do conhecimento comum, para interpretação de visão total, a fórmula:

$$(CC_g)^{\varepsilon}\varphi \equiv \varphi \wedge O^{\varepsilon}(CT_g)\varphi \wedge O^{\varepsilon}(O^{\varepsilon}(CT_g)\varphi) \wedge \ldots \wedge (O^{\varepsilon}(CT_g)\varphi)^{n}\varphi \wedge \ldots$$

é válida. No entanto, a fórmula correspondente envolvendo o

conhecimento comum eventual não é válida. $(CC_G)^{\Diamond}\varphi$ implica na conjunção infinita mas não conversamente. Isto porque se um número infinito de fatos é verdade eventualmente, então não necessariamente todos os fatos serão eventualmente verdadeiros simultaneamente.

Assim, confrontando as três noções de conhecimento comum, temos que o conhecimento comum corresponde a simultaneidade dos eventos, o ε-conhecimento comum a eventos que ocorrem dentro de um intervalo de tempo de tamanho ε enquanto que o conhecimento comum eventual correspondem a eventos que vão ocorrer em todos os pontos eventualmente. Logo, em canais de comunicação assíncrona, o fato f = "envio de m" é de $\Diamond$-conhecimento comum no momento em que m é enviado.

Por outro lado, considere uma interpretação particular I em que um processo, que recebe uma mensagem enviada em ε unidades de tempo, suporta imediatamente $(CC_G)f$ (onde f = "envio de m"). Como as ações dos processos são tidas como baseadas no seu conhecimento, I não será necessariamente uma interpretação de conhecimento já que ela pode não ter o conhecimento consistente. Como exemplo, basta considerar um processo que tenha o conhecimento de que outro processo tenha o conhecimento de f, quando de fato o outro processo ainda não o tem. Isto só acontecerá após ε unidades de tempo em que o processo passou a suportar $(CC_G)f$ pois, foi neste instante, que o último processo recebeu a mensagem m.

Voltando ao processo de comunicação, é interessante ter a noção de como os estados de conhecimento $(CC_G)^{\varepsilon}$ e $(CC_G)^{\Diamond}$ são afetados quando a comunicação não é garantida. Pelo teorema (5.3.1) e lema (5.3.1), se a comunicação não

for garantida, independente do seu tipo (síncrona ou assíncrona), o conhecimento comum não é atingido. No entanto, para o ε-conhecimento comum (conhecimento comum eventual), se $(CC_G)^{\varepsilon}\varphi$ ($(CC_G)^{\diamondsuit}\varphi$) for indeterminada no ponto (r,t), na ausência de qualquer comunicação futura, $(CC_G)^{\varepsilon}\varphi$ ($(CC_G)^{\diamondsuit}\varphi$) poderá ser verdadeira , como mostra o exemplo:

Exemplo 5.4.2 (MOSES, 1986): Considere um sistema constituido por dois processos p_1 e p_2, conectados por uma rede de comunicação, com *clocks* perfeitamente sincronizados, onde a comunicação não é garantida. Seja P o seguinte protocolo: *"no tempo t = 0, envie a mensagem "OK". Para todo t > 0, se foram recebidas t mensagens de "OK" no tempo t do clock, então envie uma mensagem de "OK" no tempo t. Caso contrário não envie nenhuma mensagem"*. Seja φ um fato indicando que **nenhuma mensagem foi entregue no intervalo de uma unidade de tempo.** Faça ε = 1. Logo, $(CC_G)^{\varepsilon}\varphi$ será indeterminada no tempo t = 0 (de acordo com o protocolo, em t = 0 foi enviada uma mensagem), pois, no melhor caso, quando todas as mensagens forem entregues dentro de uma unidade de tempo, φ nunca será verdadeira e consequentemente $(CC_G)^{\varepsilon}\varphi$ nunca será verdadeira também. Por outro lado, se φ for verdadeira em algum momento, automaticamente, $(CC_G)^{\varepsilon}\varphi$ também o será (Para isto, basta o processo p_2 falhar ao receber uma mensagem de "OK" entre um tempo t-1 e t. Com isto, p_2 não enviará uma mensagem de "OK" para p_1 e consequentemente, p_1 terá o conhecimento de φ no tempo t+1). Logo, $\varphi \supset O^{\varepsilon}(CT_G)\varphi$ é garantido e, pelo axioma da indução, $\varphi \supset (CC_G)^{\varepsilon}\varphi$ é verdadeira (o mesmo exemplo pode ser aplicado a $(CC_G)^{\diamondsuit}\varphi$). ■

O teorema seguinte, análogo ao teorema (5.3.1), prova que, em sistemas em que a comunicação não é garantida, não é possível que uma comunicação com sucesso valide $(CC_G)^{\varepsilon}\varphi$ ou $(CC_G)^{\diamondsuit}\varphi$.

Teorema 5.4.2 (MOSES, 1986): Seja φ um fato estável e S um sistema em que a comunicação não é garantida. Seja r_1, $r_2 \in$ S *runs* tais que r_2 estende (r_1,t_1) e nenhuma mensagem é entregue em r_2 para todo $t \geq t_1$. Se $(r_2,t) \models \neg(CC_G)^{\varepsilon}\varphi$ (respec. $(r_2,t) \models \neg(CC_G)^{\diamondsuit}\varphi$) então $(r_1,t) \models \neg(CC_G)^{\varepsilon}\varphi$ (respec. $(r_1,t) \models \neg(CC_G)^{\diamondsuit}\varphi$) para todo $t \geq t_1$. ■

prova: Apêndice B ■

Pelo teorema (5.4.2), mesmo havendo em r_1 uma comunicação com sucesso, tem-se que $(CC_G)^{\varepsilon}\varphi$ $((CC_G)^{\diamondsuit}\varphi)$ não será atingido. Este teorema mostra então que a comunicação não confiável não pode ser usada para planejamento em ação coordenada que garanta a participação de todos os agentes. Dai o seguinte corolário para o caso do ataque coordenado:

Corolário 5.4.1: No problema do ataque coordenado, qualquer protocolo que garanta que se uma divisão atacar, a outra divisão eventualmente atacará, é um protocolo em que nenhuma das divisões atacam. ■

prova: Apêndice B ■

Conhecimento comum provável

Uma outra variação do conhecimento comum que será tratada, estará relacionada a SD em que a comunicação não é garantida, existindo incertezas na entrega da mensagem. Com isto, a toda mensagem enviada está relacionada uma probabilidade da mesma ser recebida, ou seja, a comunicação

entre os processos é um fato **provável** de acontecer.

Considere então sistemas nos quais a entrega da mensagem, quando ocorre com sucesso é imediata. Logo, dá-se o nome de **Conhecimento comum provável** ao maior ponto fixo da equação:
$(CC_G)^P\varphi \equiv \varphi \wedge (CT_G)^P(CC_G)^P\varphi$, onde $(CT_G)^P\varphi$ significa "Provavelmente $(CT_G)\varphi$".

Por definição, $(CC_G)^P$ satisfaz o axioma do ponto fixo, já anteriormente citado. Como a noção de probabilidade não satisfaz o axioma do fecho sob a consequência, é de se esperar que $(CC_G)^P$ não satisfaça os outros dois axiomas também já definidos. Com relação ao axioma da indução, $(CC_G)^P$ satisfaz a uma noção mais fraca:

$$CC_G(\varphi \wedge (CT_G)^P\varphi) \supset (\varphi \wedge (CC_G)^P\varphi)$$

Este axioma afirma que se é de conhecimento comum que a mensagem será provavelmente entregue a todos os processos, então, ao receber a mensagem, o processo conclui que o envio da mesma é de conhecimento comum provável.

No problema do ataque coordenado, é possível ter o conhecimento comum provável da mensagem enviada do general A para o general B "dentro de uma hora". Para isto, basta que seja de conhecimento comum que provavelmente o mensageiro irá entregar a mensagem.

Outros exemplos de conhecimento comum provável têm referências a termos que possuem mais de um significado: Ao mencionar um termo com duplo significado, provavelmente os membros da comunidade irão relacionar o conhecimento a um

significado, muito embora, um elemento que não esteja totalmente integrado no contexto, possa relacionar a um outro significado que o termo venha a ter. Logo, o conhecimento provável seria mais adequado a esta situação. Assim, "as pessoas", ao acreditarem que estão atingindo o conhecimento comum, podem na realidade estar atingindo um conhecimento comum provável, o qual pode estar bem próximo da certeza.

O que resta é relacionar um grau de probabilidade, que, ao ser quantificado, pode gerar novas variantes do conhecimento comum provável através do maior ponto fixo da equação correspondente. Estas variações seriam relativas ao conhecimento comum provável com "probabilidade 1", "probabilidade x", etc.. Para cada uma destas noções, é possível caracterizar as propriedades dos diversos sistemas de comunicação e trabalhar em diversos níveis de abstração.

seção 5.5

CONHECIMENTO X AÇÃO X COMUNICAÇÃO

O quebra-cabeça dos **"maridos infiéis"** ilustra a interdependência entre conhecimento e ação levando em consideração os tipos de canal de comunicação ao qual o sistema está submetido.

Esta seção tem por objetivo interagir as noções de conhecimento, ação e comunicação e analisar os problemas existentes nesta interação. Esta interação é importante quando a comunicação é utilizada para transmitir a informação (mensagem), a qual pode definir ações subsequentes. Em SD, os problemas de interação dizem

respeito ao "conhecimento" que um processo pode obter apenas "observando" as "ações" tomadas por outros processos, as quais se relacionam a um fato que é de conhecimento comum.

O problema

O exemplo ilustrativo, o quebra-cabeça dos **"maridos infiéis"**, mostra a interdependência existente entre o conhecimento e a ação. Seu passo inicial consiste de um fato anunciado publicamente, o qual se torna de conhecimento comum. Baseado neste fato, é possível chegar a solução do problema considerando apenas ações tomadas pelas "esposas" sem que haja nenhuma comunicação entre as mesmas. Partindo do problema original, a cada variação, tem-se uma evolução dos processos de comunicação entre as "esposas" afetadas pelo problema dos "maridos infiéis". Em resumo, a descrição original é a seguinte (MOSES, DOLEV e HALPERN, 1985):

Existia uma cidade chamada de "Mamajorca", onde as rainhas vinham fazendo campanha contra o problema da infidelidade masculina. Considere a época de uma rainha chamada Henrietta I. Nesta época, as mulheres em Mamajorca, para terem um marido precisavam ser aprovadas em um exame de capacidade de raciocínio lógico além de ter saúde perfeita. Por outro lado, as rainhas não precisavam mostrar tamanha competência.

Os fatos de conhecimento comum em Mamajorca eram:

1. As rainhas eram pessoas perfeitamente confiáveis

2. As mulheres sempre obedeciam a rainha

3. Todas as mulheres eram capazes de ouvir qualquer disparo (tiro) que viesse a ser dado em Mamajorca.

A rainha Henrietta I ,a fim de acabar com o problema da infidelidade masculina, chamou todas as mulheres casadas, juntou-as na praça da cidade e leu o seguinte documento:

" *Existe no mínimo um marido infiel nesta comunidade. Embora nenhuma de vocês antes desta reunião tenha o conhecimento se seu marido é ou não infiel, cada uma tem o conhecimento sobre a fidelidade dos outros maridos. Todas estão proibidas de discutir este assunto com qualquer outra pessoa. Contudo, se alguma de vocês concluir que seu próprio marido é infiel, esta deve atirar no mesmo à meia-noite do dia da descoberta*".

Após esta declaração, trinta e nove noites passaram, e, na quadragésima noite, tiros foram ouvidos. O problema não expõe explicitamente quantos maridos infiéis sofreram o tiro e quantos maridos infiéis existiam em Mamajorca na época e,nem mesmo, como as esposas traídas concluiram sobre a infidelidade de seus maridos após trinta e nove noites de silêncio ou se qualquer outro marido sofreu algum tiro em alguma noite seguinte.

Para tratar destas questões, considere inicialmente a existência de apenas um marido infiel. Sua esposa, após ouvir a leitura do documento, não conhecendo outro marido

infiel, conclui sobre a infidelidade de seu marido e atira na primeira noite. Ao considerar dois maridos infiéis, suas esposas têm o conhecimento de apenas um marido infiel e esperam ouvir um tiro na primeira noite. Como isto não ocorre, elas concluem sobre a infidelidade de seus maridos e atiram na segunda noite. Baseando-se nestes argumentos, a conclusão do problema para um número n de maridos infiéis é a seguinte:

Teorema 5.5.1 (MOSES, DOLEV e HALPERN, 1985): Se existem n maridos infiéis em Mamajorca, no instante em que a rainha Henrietta I leu o documento, então as esposas traidas atirarão em seus maridos na noite do n-ésimo dia. ■

prova: Apêndice B ■

Um fato a observar é que, ao receber o comunicado da rainha, a esposa que tem o conhecimento de k maridos infiéis, tem o conhecimento de que as esposas traidas têm o conhecimento de k-1 maridos infiéis cujas esposas traidas têm o conhecimento de k-2 maridos infiéis cujas esposas traidas têm o conhecimento de k-3 maridos infiéis e assim sucessivamente até chegar ao ponto em que apenas uma esposa traida não tenha o conhecimento de outro marido infiel que não seja o seu. Isto indica que a existência de k, $k > 1$, maridos infiéis não é de conhecimento comum quando do anúncio da rainha. O conhecimento comum é obtido de acordo com as ações das esposas: uma vez que a primeira noite passou-se em silêncio, foi adquirido o conhecimento comum de no mínimo dois maridos infiéis, com a segunda noite de silêncio, torna-se de conhecimento comum a existência de no mínimo três maridos infiéis, e, ao continuar com o

processo, o conhecimento comum de k+1 maridos infiéis é obtido após a k-ésima noite de silêncio. O importante é que este conhecimento comum foi obtido sem nenhuma comunicação entre as esposas.

Aplicando pequenas variações no problema original, é possível analisar casos em que o canal de comunicacão é assíncrono, considerar os diversos tipos de comunicação síncrona e discutir as condições sob as quais o protocolo utilizado pode tolerar falhas.

a) Comunicação assíncrona

Para analisar a interdependência entre conhecimento e ação, em sistemas onde a comunicação é assíncrona, considere a seguinte variação do problema original:

Suponha que Henrietta II, sucessora de Henrietta I, para continuar combatendo o mesmo problema, implantou um sistema de correios para evitar a reunião na praça, garantindo que todas as cartas enviadas de sua corte chegariam **eventualmente** em todas as casas de Mamajorca. A primeira carta tratava deste sistema de correios. A segunda carta enviada era uma cópia fiel do documento de Henrietta I.

Conclusão: Esta segunda idéia foi um fracasso. A explicação para o fracasso se encontra na característica assíncrona do sistema de correios implantado, no qual as mensagens são entregues eventualmente e, desta forma, as esposas traidas não terão o conhecimento de que as outras esposas já receberam ou não a carta. O seguinte teorema garante o fracasso de Henrietta II:

Teorema 5.5.2 (MOSES, DOLEV e HALPERN, 1985): Se existe mais de um marido infiel, então, ao usar um canal assíncrono para a emissão das instruções, nenhum marido infiel sofrerá um tiro. ■

Quando existe apenas um marido infiel, sua esposa, ao receber a carta, irá atirar a meia-noite do mesmo dia. O problema aparece no caso de k > 1 maridos infiéis. Neste caso, não haverá tiros porque as k esposas traidas imaginarão sempre que seus maridos são fiéis e que as k-1 esposas traidas que elas conhecem ainda não receberam a carta da rainha.

Sendo a carta da rainha de conhecimento comum eventual, uma esposa nunca poderá determinar se as noites em silêncio são resultantes da reação das outras esposas ao receber a carta ou do fato delas ainda não terem recebido a carta. Assim, mesmo que todas as cartas sejam entregues simultaneamente, o fato de ser de conhecimento comum que as cartas são entregues eventualmente faz com que não seja possível descobrir os maridos infiéis.

b) Comunicação síncrona

Para evitar os problemas sofridos por Henrietta II, Henrietta III melhora o sistema de correios de modo que seja de conhecimento comum que qualquer carta, enviada pela rainha, chega a casa de suas súditas em, no máximo. um dia. Assim, a primeira carta enviada tratava desta melhoria do sistema e a segunda carta enviada foi uma cópia fiel da carta de Henrietta I.

Conclusão: Apesar desta idéia ter sido mais eficiente que a anterior, ela não chegou a ser tão boa quanto a idéia

inicial. Caso Henrietta III tivesse falado as súditas para esperar uns dias antes de atirar, ela teria atingido a mesma fama de Henrietta I. Outro detalhe é a não existência de um calendário nesta época.

Para se chegar a tais conclusões, considere $(CT_G)^{m+1}\varphi = (CT_G)(CT_G)^m\varphi$, para $m > 0$, onde $CT_G(\varphi)$ está definido na seção (2.7). O que se afirma neste caso é que, se existem n maridos infiéis, e se φ = "a rainha enviou a carta", então $(CT_G)^n\varphi$ se torna verdadeiro em algum momento, de modo que o primeiro tiro será ouvido no máximo n dias após $(CT_G)^n\varphi$ ser válido. Com isto, ao considerar que a carta será entregue a cada esposa em no máximo d dias, seu conteúdo será de d-conhecimento comum (dentro de d dias todas recebem a carta e dentro de mais d dias todas sabem que todas receberam a carta e assim por diante). Logo, kd dias após a rainha enviar a carta, $(CT_G)^k\varphi$ é válido e no mínimo um tiro será ouvido. Este raciocínio está formalizado na proposição (5.5.1):

Seja d o número máximo de dias que a carta da rainha leva para ser entregue as suas súditas

Proposição 5.5.1 (MOSES, DOLEV e HALPERN, 1985): No caso síncrono com limite de entrega de d dias,uma esposa que conhece k maridos infiéis, terá o conhecimento de que seu próprio marido é infiel se kd noites em silêncio passarem após o dia no qual ela recebeu a carta da rainha. ■

O exemplo seguinte justifica porque Henrietta III não foi tão bem sucedida quanto Henrietta I, podendo ter

cometido injustiças:

Exemplo 5.5.1: Considere um grupo g composto de duas esposas p_1 e p_2 e o dia d = 2. Suponha que p_1 tenha o conhecimento de que o marido de p_2 é infiel, e que p_1 recebeu a carta da rainha na segunda-feira. À meia noite da terça-feira, p_2 atirou em seu marido. Devido a ausência de um calendário, p_1 não é capaz de concluir sobre a fidelidade de seu marido, já que duas hipóteses são possíveis de acontecer:

1. O marido de p_1 é fiel e p_2 recebeu a carta na terça-feira, fazendo com que seu marido sofresse o tiro na terça-feira a noite.
2. O marido de p_1 é infiel e p_2 recebeu a carta no domingo. Com isto, esperou o domingo e a segunda-feira para ouvir um tiro de p_1 e, como não aconteceu, p_2 concluiu que seu marido era infiel e atirou na noite de terça-feira. ■

Este exemplo mostra que p_1 permanecerá na duvida quanto a fidelidade de seu marido se considerar apenas as ações de p_2.

Chamando de **1º dia significativo** o primeiro dia em que uma esposa traida recebe a carta , é fácil ver que as esposas traidas que recebem a carta no 1º dia significativo serão as primeiras a atirar em seus maridos. No entanto as outras esposas traidas ficarão na dúvida quanto a fidelidade de seus maridos, como pode ser visto pelo seguinte teorema:

Teorema 5.5.3 (MOSES, DOLEV e HALPERN, 1985): Usando um canal de comunicação síncrono, as esposas traidas que receberam a carta da rainha no 1º dia significativo

atirarão em seus maridos (n-1)d dias após esse dia, onde n é o número de maridos infiéis. Todas as outras esposas permanecerão na dúvida quanto a fidelidade de seus maridos.
prova: Apêndice B ■

Para solucionar o problema da dúvida, é sugerido que as esposas traídas, ao descobrirem sobre a infidelidade de seus maridos, façam uso de um certo tempo de espera, após descobrir sobre a infidelidade, antes de atirarem:

Proposição 5.5.2 (MOSES, DOLEV e HALPERN, 1985): Em sistemas onde a comunicação é síncrona com limite de entrega de d dias, se toda esposa esperar e dias a partir do dia em que ela descobre sobre a infidelidade de seu marido, então a esposa que tem o conhecimento de k maridos infiéis, saberá que seu próprio marido é infiel se k(d+e) noites silenciosas passarem a partir do dia em que ela recebeu a carta. ■
Esboço da prova: A prova se dá por indução no número de maridos infiéis. ■

O seguinte teorema certifica esta sugestão:

Teorema 5.5.4 (MOSES, DOLEV e HALPERN, 1985): Se a espera é longa o suficiente, ou seja, $e \geq d-1$, então a esposa traída atira em seu marido sem deixar dúvidas quanto ao problema da infidelidade. ■
Esboço da prova: Semelhante a prova do teorema (5.5.3) ■

Outro caso a considerar é a existência de suborno: uma

esposa pode subornar o carteiro para saber exatamente o dia em que sua carta foi enviada pela rainha; refletindo o caso em que o protocolo utilizado pudesse tolerar falhas. Seria então de conhecimento comum:

i. O fato do suborno ser um segredo entre a esposa e o carteiro subornado e,

ii. O fato de nenhuma esposa ter o conhecimento das outras esposas que subornaram o carteiro.

Neste caso, a seguinte proposição é verdadeira:

Proposição 5.5.3 (MOSES, DOLEV e HALPERN, 1985): No caso da comunicação síncrona, a esposa que suborna o carteiro para saber quando a rainha enviou originalmente a carta, eventualmente terá o conhecimento se seu marido é fiel. ■

Esboço da prova: A prova se relaciona ao fato de, com o suborno, a esposa ter o conhecimento de quando se passaram kd noites e assim ter o conhecimento eventual sobre a fidelidade de seu marido. ■

c) caso fortemente síncrono

Os problemas vistos acima se devem ao fato da comunicação não ser fortemente assíncrona. Portanto, se no reino de Henrietta III fosse implantado um calendário, todas as esposas teriam o conhecimento comum da data do envio da carta da rainha. Analisemos o que ocorreria neste caso:

No reinado de Henrietta IV, foi implantado um calendário. Para comunicar sobre o calendário, Henrietta IV reuniu-se com todas as esposas na praça e anunciou que a

partir daquele momento, o sistema de correios seria fortemente síncrono, isto é, toda carta enviada pela rainha conterá a data de expedição e tem a garantia de ser entregue a todas em no máximo d dias.

Henrietta IV enviou uma carta a todas as esposas datada do dia de expedição e contendo as instruções de Henrietta I. Passou-se o tempo e nada aconteceu. Com isto Henrietta IV concluiu que com o procedimento de Henrietta III (situação anterior), a confiança presente na mornaquia tinha sido perdida, e algumas esposas já não obedeciam mais as ordens vindas da rainha. Henrietta IV enviou uma nova carta contendo a seguinte frase:
" *Existe no mínimo uma esposa obediente cujo marido é infiel*"

Assim, a confiança voltou ao reino e a seguinte proposição diz porque a segunda carta foi tão importante para recuperar esta confiança:

Proposição 5.5.4 (MOSES, DOLEV e HALPERN, 1985): No caso fortemente síncrono, se existe exatamente uma esposa traida e ela é desobediente, todas as outras esposas estão sujeitas a atirar em seus maridos na segunda noite. ■
Esboço da prova: Se as outras esposas não sabem da desobediência da esposa traida, na segunda noite todos os maridos serão eliminados e o infiel sobrevive. ■

Mesmo que todas as esposas sejam obedientes, sem a segunda carta, elas não atirarão: considere que existam duas esposas traidas no grupo. Caso a última carta de

Henrietta IV não tivesse sido enviada, no segundo dia cada esposa traida não iria atirar pois não saberia se o silêncio era devido a infidelidade de seu marido ou devido a desobediência da esposa traida. Assim, a partir da segunda noite, não seriam ouvidos mais tiros. No caso geral, argumenta-se por indução no número de esposas traidas.

O seguinte teorema mostra como a segunda carta é importante:

Teorema 5.5.5 (MOSES, DOLEV e HALPERN, 1985): no caso fortemente síncrono, se é de conhecimento comum que existe no mínimo uma esposa traida que é obediente, então todas as esposas traidas obedientes irão atirar em seus maridos. ■

A diferença entre o caso de suborno e o caso fortemente síncrono é que no primeiro se todas as esposas subornarem, elas têm o conhecimento do dia em que a rainha enviou a carta, mas nenhuma delas tem o conhecimento sobre o conhecimento das outras. Já no segundo caso, o dia em que a rainha enviou a carta é de conhecimento comum. Apesar do efeito ser o mesmo, nos dois casos todos os maridos infiéis são eliminados e nenhuma dúvida quanto a fidelidade existe. No fim do processo, as esposas levam mais tempo para eliminarem seus maridos no 1º caso do que no 2º. Também no 1º caso, enquanto os maridos não são eliminados, as esposas não sabem se alguma injustiça será cometida enquanto no 2º caso, é de conhecimento comum que haverá justiça.

círculo de comunicação

Uma outra abordagem para tratar o problema da infidelidade é sugerida quando se tem um posicionamento das "residências" de modo a formar um círculo. Em resumo, ter-se-ia a seguinte descrição:

Seria de conhecimento comum o posicionamento das residências bem como o fato das cartas serem entregues segundo o sentido dos ponteiros do relógio. As rainhas tentaram resolver o problema enviando a cópia da carta de Henrietta II e utilizando um sistema de correios com as mesmas características do sistema de Mamajorca referente as respectivas épocas. A conclusão foi que neste caso, nenhuma rainha chegou a fracassar como Henrietta II, mas também nenhuma chegou a ter o êxito de Henrietta IV. O seguinte teorema justifica estas conclusões:

Teorema 5.5.6 (MOSES, DOLEV e HALPERN, 1985): Ao obedecer a uma organização em círculos, os seguintes resultados são obtidos:

1. Quando a comunicação é assíncrona, a última esposa traida a receber a carta irá atirar em seu marido enquanto nenhuma outra o fará.
2. Quando a comunicação é síncrona, algumas esposas traidas atiram em seus maridos enquanto outras não o fazem.
3. Quando a comunicação é fortemente síncrona, algumas esposas traidas atiram em seus maridos enquanto outras não o fazem. ■

Esboço da prova:

1. A prova é dada por indução no número de maridos infiéis. Esta prova baseia-se no fato de que a última esposa traida

tem o conhecimento de que todas as outras já receberam a carta e não atiraram devido a natureza assíncrona do processo de comunicação, enquanto ela, por ser a última, pode com certeza deduzir sobre a infidelidade de seu marido.

2. Para mostrar que algumas esposas atirarão em seus maridos, basta fazer indução no número de maridos infiéis. Para mostrar que a injustiça ocorre, apela-se para o seguinte exemplo: considere duas esposas traidas, p_1 e p_2, $d = 2$ e suponha que p_2 tem o conhecimento de que o marido de p_1 é infiel. Suponha também que p_2 recebeu a carta no domingo e que p_1 atirou em seu marido na segunda-feira. p_2 então não conseguirá distinguir entre estas duas possibilidades:

2.1. p_1 recebeu a carta no domingo e, sabendo que o marido de p_2 é infiel, ela esperou o tiro no domingo a noite. como isto não aconteceu, p_1 atirou em seu marido na segunda-feira a noite.

2.2. p_1 recebeu a carta na segunda-feira e sabendo que o marido de p_2 é fiel, concluiu sobre a infidelidade de seu marido e atirou na mesma noite. Logo, o caso (2.1) seria uma injustiça.

3. A prova é semelhante ao caso (2), apenas considerando o dia oficial do envio da carta como sendo o domingo. ■

Pelo teorema, no caso (1), o conhecimento da ordem de entrega ajuda a última esposa traida a descobrir sobre a infidelidade de seu marido, ou seja, o conhecimento da ordem de entrega da mensagem é útil. Já no caso (3), esta ordem contribui negativamente visto que, se ela não

existisse, o problema seria resolvido com sucesso.

Seção 5.6

CONCLUSÕES

Os dois problemas são considerados com os seguintes objetivos: utiliza-se o primeiro ("Ataque coordenado") para mostrar que se a comunicação não for garantida, é impossível usá-la para coordenar uma ação simultânea. Já o segundo ("esposas traidas e maridos infiéis") trata não só da interação entre conhecimento, ação e comunicação, mas suas variantes consideram as propriedades dos diversos tipos de comunicação. O importante, neste caso, é mostrar que o conhecimento é obtido através das observações dos elementos do sistema (processos), conhecendo apenas suas ações descritas através dos fatos que são de conhecimento comum. Neste problema, o protocolo a ser seguido é formado pelas instruções da rainha.

É válido observar que as idéias apresentadas em (MOSES, 1986), (HALPERN e MOSES, 1984), (MOSES, DOLEV e HALPERN, 1985) e (HALPERN, 1985) mostram a sutileza no relacionamento entre conhecimento, ação e comunicação de modo bem geral. Resta tratar do conhecimento em SD considerando suas propriedades particulares além de utilizá-lo para analisar protocolos particulares.

Observe também que a estrutura utilizada só considera o conhecimento de fatos verdadeiros. Quanto as formalizações, não foi analisada aqui as possibilidades utilizando a noção de crença ou incorporando a noção de probabilidade. Como o modelo de SD apresentado só obedece

ao sistema modal S5, podem existir outras propriedades específicas de um SD que, ao serem consideradas, poderão causar grandes mudanças quando se trabalha com este tipo de problema. O que MOSES afirma é que sua ferramenta é geral o suficiente para acomodar qualquer noção simples de conhecimento.

CAPÍTULO VI

CONCLUSÃO

Neste trabalho foram estudadas algumas abordagens lógicas para formalizar a noção de conhecimento e crença.

Inicialmente foi definido um modelo geral para o conhecimento onde várias lógicas modais foram expressas, as quais permitiam uma boa aproximação para o "raciocínio" de uma base de conhecimento. A noção de crença que também foi definida pode ser interpretada como uma versão mais "fraca" do conhecimento. O modelo semântico utilizado foi a estrutura *Kripke*, muito embora outros modelos, como a estrutura de conhecimento, possam ser utilizados de acordo com a aplicação desejada.

Ao utilizar o modelo dos mundos possíveis, não é possível formalizar alguns aspectos do "raciocínio" humano visto que os agentes, neste modelo, sofrem de um problema chamado omnisciência lógica. Este problema, bem como as suas causas, nos levou a analisar várias abordagens para resolvê-lo. As abordagens estudadas foram: a lógica da crença implícita e explícita, a lógica da consciência, a lógica da consciência geral, a lógica do raciocínio local e a lógica proposicional não-padrão. É difícil dizer qual destas abordagens é a mais poderosa. Com certeza, nenhuma delas é suficientemente geral para esta formalização. Porém, a depender da aplicação, uma abordagem "adequada" poderá ser encontrada.

Uma característica presente na maioria destas abordagens diz respeito a monotonicidade, ou seja, ao

termos uma base de conhecimento contendo fatos verdadeiros, adicionar novos fatos a esta base não irá invalidar os fatos já existentes. O "raciocínio" humano por sua vez se aproxima de um "raciocínio" não-monotônico e, por esta razão, apresentamos alguns formalismos baseados na lógica do conhecimento e da crença nos quais, ao contrário dos anteriores, o processo de dedução está baseado não apenas na informação que o agente possui, mas no fato desta informação ser considerada todo o conhecimento (ou crença) do agente.

Foi apresentada também uma aplicação na especificação de Sistemas Distribuídos. Dois problemas foram analisados, os quais determinam uma interação entre conhecimento, ação e comunicação. Entre outros resultados, foi mostrado que se a comunicação não é garantida ou confiável, não é possível utilizá-la para coordenar uma ação simultânea. Mostrou-se também a "influência" do conhecimento na relação entre a ação e a comunicação. Na realidade, para relacionar o conhecimento e a ação em um sistema distribuido, é preciso determinar os estados de conhecimento dos processos necessários a execução de uma determinada tarefa e relacionar estes estados aos efeitos trazidos pelas propriedades deste sistema.

Este trabalho tem sua importância pois constitui um estudo introdutório e específico sobre a lógica do conhecimento ou crença. É direcionado àqueles interessados em temas como formalização do conhecimento, raciocínio não-monotônico, especificação de sistemas distribuídos, bem como em aplicações da lógica do conhecimento nos diversos campos da ciência da computação. Citamos então algumas

sugestões para pesquisas futuras:

i) Analisar os problemas relacionados a complexidade de se determinar a satisfatibilidade das fórmulas nos formalismos apresentados.

ii) Analisar e desenvolver versões quantificadas das formalizações lógicas aqui introduzidas, cujo objetivo seria descrever situações mais complexas.

iii) Especificamente, na abordagen de HALPERN e MOSES (1984) apresentada no capítulo IV, Analisar os problemas existentes ao levantar a possibilidade de uma extensão para mais de um agente. Quanto a abordagem de LEVESQUE (1990) um dos problemas deixados diz respeito a prova de completude para o caso quantificacional.

iv) Uma outra linha de pesquisa, seria acrescentar aos modelos aqui definidos uma relação entre a noção de conhecimento, crença e probabilidade, onde, por exemplo, poderíamos ter fatos como: "de acordo com o agente i, a fórmula φ é verdadeira com probabilidade de no mínimo α". Neste sentido, podemos citar como referências (FAGIN e HALPERN, 1988a), (FAGIN e HALPERN, 1988b) e (HALPERN, 1989).

iv) Além destes, na análise de sistemas distribuídos, pode-se dar um tratamento aos sistemas distribuídos práticos, os quais não foram considerados, onde as

incertezas relativas a leitura dos *clocks* ou no tempo de transmissão da mensagem impossibilitam uma modificação simultânea na visão dos processos.

APÊNDICE A

Teorema 2.5.1 (HALPERN e MOSES, 1985): K_m é uma axiomatização correta e completa para mundos Kripke. ■

A prova deste teorema utiliza alguns conceitos que são apresentados a seguir:

Define-se uma fórmula α como sendo uma **fórmula consistente** com um sistema de axiomas se $\neg\alpha$ não for provada deste sistema. Por sua vez, α é **provada** de um sistema de axiomas se α for uma instância de um dos axiomas do sistema ou se for derivada de fórmulas provadas através de uma das regras de inferência do sistema. Define-se também um conjunto F de fórmulas como sendo um **conjunto maximal consistente** se ele é consistente e se, para toda fórmula $\alpha \in L_m$ que não está contida em F, $F \cup \{\alpha\}$ não é consistente.

Além desses, usando as técnicas padrões do raciocínio proposicional, mostra-se que (HALPERN e MOSES, 1985):

Lema 2.5.1: Em qualquer sistema axiomático que contenha A.1 e R.1, temos que:

1. Todo conjunto consistente F pode ser estendido a um conjunto maximal consistente.

2. Se F é um conjunto maximal consistente, então para todas as fórmulas p e q:

   i) $p \in F$ ou $\neg p \in F$

   ii) $p \wedge q \in F$ sse $p \in F$ e $q \in F$

   iii) se $p \in F$ e $p \supset q \in F$ então $q \in F$

   iv) se p é uma fórmula válida então $p \in F$

■

Esboço da prova do teorema 2.5.1: As propriedades de $\models$ já descritas implicam que K_m é correto com respeito aos mundos Kripke. Para provar a completude, basta provar que toda fórmula consistente é satisfatível. A prova é então realizada da seguinte forma:

Constrói-se uma estrutura Kripke canônica K_c para o sistema K_m, contendo um estado s_f para cada conjunto maximal consistente F de K_m, de modo que para toda fórmula $\alpha \in F$, $(K_c, s_f) \models \alpha$:

Define-se inicialmente $F/C_i = \{\alpha \mid C_i\alpha \in F\}$, para cada conjunto maximal consistente F. $K_c = (S, \pi, \rho_1, \ldots, \rho_n)$ é então definido da seguinte forma:

- $S = \{s_f \mid F$ é um conjunto maximal consistente$\}$

- $\pi(s_f, p)$ será verdadeiro se $p \in F$ e será falso em caso contrário, onde p é uma proposição primitiva.

- Para cada $i = 1, \ldots, n$, $(s_f, s_v) \in \rho_i$ sse $F/C_i \subseteq W$ onde F e W são conjuntos maximais consistentes.

Como todo conjunto de fórmulas consistente está contido em algum conjunto maximal consistente (Lema 2.5.1), mostra-se então por indução na estrutura de α que para todo conjunto maximal consistente F, $(K_c, s_f) \models \alpha$ sse $\alpha \in F$. ■

Para os teoremas seguintes, como as provas de "correção" e "completude" seguem um raciocínio bem semelhante a prova do teorema (2.5.1), será deixado apenas a estrutura Kripke canônica para cada uma das lógicas consideradas.

Teorema 3.3.1 (HALPERN e FAGIN, 1988): Acrescentando A.14 aos axiomas de KD45m, obtém-se uma axiomatização correta e completa para a lógica da consciência. ■

Esboço da prova: A prova é semelhante a prova do teorema (2.5.1). Neste caso, a estrutura Kripke canônica K_c tem a seguinte construção: $K_c = (S, \pi, A_1, \ldots, A_m, B_1, \ldots, B_n)$ onde,

1. S e π são definidos como no teorema (2.5.1)

$A_i(s_f) = \{p \mid B_i(p \vee \neg p) \in F\}$

$B_i = \{(s_f, s_v) \mid F/L_i \subseteq W\}$ ■

Teorema 3.4.1 (HALPERN e FAGIN, 1988): Acrescentando o axioma A.15 aos axiomas de KD45m, obtém-se uma axiomatização correta e completa para a lógica da consciência geral. ■

Esboço da prova: Neste caso, a estrutura Kripke construida é idêntica a estrutura construida no teorema (3.3.1), com a diferença que agora $A_i(s_F) = \{\varphi \mid A_i\varphi \in s_F\}$. ■

Teorema 3.5.1 (HALPERN e FAGIN, 1988): Os axiomas A.1, A.2 (cap.2), A.16 e A.17 , e as regras de inferência R.1 (cap.2), R.4 e R.5 formam uma axiomatização correta e completa para a lógica do raciocínio local. ■

Esboço da prova: Neste caso, a estrutura Kripke canônica terá dois estados correspondente a cada conjunto maximal consistente. Isto permitirá tratar o conhecimento implícito diretamente. A estrutura $K_c = (S, \pi, C_1, \ldots C_n)$ é definida da seguinte forma:

- $S = \{s_f^h$, tal que F é um conjunto maximal consistente; $h = 0,1\}$

- $\pi(s_f^h, p)$, $h = 0,1$, será verdadeiro se $p \in F$ e falso em caso contrário, onde p é uma proposição primitiva.

- $C_i(s_f^h) = \{T_{\psi,f}^{h'} \mid B_i\psi \in F;\ h = 0,1\}$ onde,

$T_{\psi,f}^{h'} = \{s_v^h \mid \psi \in W\} \cup \{s_v^l \mid F/L_i \subseteq W;\ l = 0,1\}$ onde F, W

são conjuntos maximais consistentes.

Inicialmente, para mostrar que esta estrutura é uma estrutura kripke para o raciocínio local, mostra-se que $Ci(s_f^h)$ é um conjunto não vazio de subconjuntos não vazios de S. Segue então com a prova usual. ■

Teorema 4.6.1 (RAO e FOO, 1987): O sistema modal S5Em é uma axiomatização correta e completa para mundos reflexivos-simétricos-euclidianos, cuja a função j_i satisfaz a condição C.1. ■

Esboço da prova: A estrutura kripke canônica $K_c = (S, \pi, \rho_1, \ldots, \rho_n, j_1, \ldots, j_n)$ é definida da seguinte forma:

- S e π são definidos da maneira usual
- Para $sf \in F$, se $j_i(\varphi, sf) = \langle \alpha, \beta \rangle$ então $Pi(\varphi, \alpha) \in F$ e $Ni(\varphi, \beta) \in F$.
- Para $sf \in F$, se $Pi(\varphi, \alpha) \in F$ então $j_i(\varphi, sf) = \langle \alpha, \beta \rangle$
- Para $sf \in F$, se $Ni(\varphi, \beta) \in F$ então $j_i(\varphi, sf) = \langle \alpha, \beta \rangle$
- Para $sf \in F$, $Ci\varphi \in F$, $Pi(\varphi, \alpha) \in F$, $Ci\alpha \in F$, $Ni(\varphi, \beta) \in F$, $\neg Ci\beta \in F$ sse $Ei(\varphi, \alpha, \beta) \in F$
- Para qualquer proposição primitiva p e qualquer $sf \in F$, $(K_c, sf) \models \varphi$ se $\varphi \in F$, caso contrário $(K_c, sf) \not\models \varphi$. ■

APÊNDICE B

lema 5.2.1 (MOSES, 1986): Se I é uma interpretação de conhecimento para S e r ∈ S, então, qualquer que seja $p_i \in g$, para todo ponto (r,t), p_i suporta $C_G(\varphi)$ em (I,r,t) sse $(I,r,t) \models C_G(\varphi)$. ■

prova: A direção "se" segue do fato de $C_i(\varphi) \supset \varphi$ ser obedecido e da própria definição do fato de p_i suportar $C_G(\varphi)$.

Na outra direção, proceda da seguinte forma: Suponha o contrário: que p_i suporta $C_G(\varphi)$ e $(I,r,t) \not\models C_G(\varphi)$. Logo, existe uma fórmula $\alpha = (C_{i_1})(C_{i_2})\ldots(C_{i_n})(\varphi)$ tal que $(I,r,t) \not\models \alpha$. Mas, por hipótese, se p_i suporta $C_G(\varphi)$ então $(I,r,t) \models C_i(\alpha)$. Como I é uma interpretação de conhecimento, deveríamos ter $(I,r,t) \models \alpha$, o que não nos permite ter $(I,r,t) \not\models C_G(\varphi)$. Logo, o lema é verdadeiro. ■

Lema 5.3.1 (MOSES, 1986): Seja S um sistema onde a comunicação não é confiável e seja I uma interpretação de conhecimento para S. Sejam $r, r_1 \in S$ *runs* tais que r estende (r_1,t_1) e seja $t \geq t_1$. Então para todas as fórmulas φ, $(I,r_1,t) \models C_G(\varphi)$ sse $(I,r,t) \models C_G(\varphi)$. ■

Esboço da prova : Considera-se inicialmente uma *run* r^- que estende (r_1,t_1) na qual nenhuma mensagem é entregue após o tempo t_1. Pela definição (5.3.1), $r^- \in S$. A idéia é então provar, por indução no número n de mensagens entregues em r no intervalo de t_1 a t (sem incluí-lo), que, para toda *run* r que estende (r_1,t_1), $(I,r,t) \models C_G(\varphi)$ sse $(I,r^-,t) \models C_G(\varphi)$. ■

Teorema 5.3.1 (MOSES, 1986): Seja S um sistema onde a comunicação não é garantida tal que $C^{C_G}(\varphi)$ é indeterminada em S em (r_1,t_1). Se $r \in S$ estende (r_1,t_1) e $t \geq t_1$ então, em (r,t), $C^{C_G}(\varphi)$ não é verdadeira. ■

Esboço da prova: Como $C^{C_G}(\varphi)$ é indeterminado em S em (r_1,t_1), existirá uma *run* $r' \in S$ que estende (r_1,t_1) onde $C^{C_G}(\varphi)$ não é verdadeira em (r',t') para $t' \geq t_1$. COnsidere uma *run* $r \in S$ que estende (r_1,t_1) e seja $t \geq t_1$. Pelo lema (5.3.1), $C^{C_G}(\varphi)$ não é verdadeiro em (r,t). Como r e t foram escolhidos arbitrariamente, o teorema é válido. ■

proposição 5.3.1 (MOSES, 1986): No problema do ataque coordenado, qualquer protocolo que garanta que se uma divisão atacar, então as duas divisões atacarão simultaneamente, é um protocolo em que necessariamente nenhuma das divisões irá atacar. ■

esboço da prova:

Considerações iniciais:

1. O problema por si descreve um estado específico
2. Os generais G_A e G_B são tratados como processos, cuja linha de comunicação é representada pelas mensagens trocadas entre eles.
3. A situação inicial descrita se encontra no tempo t_1.
4. É assumido que os generais obedecem a um *joint* protocolo determinítico (P_A,P_B) onde G_A segue P_A e G_B segue P_B.
5. As ações dos generais são funções determinísticas de sua história da mensagem e do tempo de seu *clock*.
6. O problema se relaciona a um SD S, onde as *runs* de S são todas as possíveis *runs* de (P_A,P_B) a partir de t_1.

Considera-se que o *joint* protocolo (P_A,P_B) garante que nenhum general irá atacar sozinho e que nenhum general irá atacar caso não haja qualquer comunicação com sucesso. Logo, tem-se que , para $p_i \in \{A,B\}$, $r \in S$ e $t \geq t_1$, existirão dois estados possíveis: $s(p_i,r,t)$ = "ATAQUE" se p_i inicia o ataque em (r,t) ou $s(p_i,r,t)$ = "NÃO ATAQUE" em caso contrário. Considera-se também uma interpretação de conhecimento baseada em estado relativa a S, I_e, onde S é um sistema onde a comunicação não é garantida e $\bar{r} \in S$ é uma *run* na qual nenhuma mensagem é entregue com sucesso e, consequentemente, nenhum general irá atacar.

Chame φ o seguinte fato: "os generais A e B estão atacando". Já que para todo $r \in S$, $\bar{r}$ estende (r,t_1), segue que φ é uma fórmula indeterminada em (r,t_1) para todo $r \in S$. Como $(CC_g)\varphi \supset \varphi$ é válido, se φ é indeterminado em (r,t_1) então para todo $r \in S$, $(CC_g)\varphi$ também o será. Como (P_A,P_B) garante o ataque coordenado, segue que em todos os pontos $(r,t) \in S \times [t_1,\infty)$ tem-se $s(G_A,r,t) = s(G_B,r,t)$. Como consequência, todos os pontos atingidos de (r,t) onde os generais estão em estado de ataque, possuem a propriedade dos dois generais se encontrarem neste estado. Logo, $(CC_g)\varphi$ é verdadeiro apenas quando os generais atacam. Por fim, pelo teorema (5.3.1), MOSES (1986) prova que como a comunicação não é garantida em S e φ é um fato inicialmente indeterminado,tem-se que φ nunca será de conhecimento comum em S. ■

Teorema 5.4.2 (MOSES, 1986): Seja φ um fato estável e S um sistema em que a comunicação não é garantida. Seja r_1, $r_2 \in S$ *runs* tais que r_2 estende (r_1,t_1) e nenhuma mensagem é

entregue em r_2 para todo $t \geq t_1$. Se $(r_2,t) \models \neg(C C_g)^{\varepsilon}\varphi$ (respec. $(r_2,t) \models \neg(C C_g)^{\Diamond}\varphi$) então $(r_1,t) \models \neg(C C_g)^{\varepsilon}\varphi$ (respec. $(r_1,t) \models \neg(C C_g)^{\Diamond}\varphi$) para todo $t \geq t_1$. ■

esboço da prova: Pode-se provar por indução em n que não existe uma *run* $r \in S$ que estenda (r_1,t_1) tal que $(C C_g)^{\varepsilon}\varphi$ ($(C C_g)^{\Diamond}\varphi$) é verdadeira em $t \geq t_1$ e exatamente n mensagens são entregues ao seu destino até o momento em que o primeiro processador toma o conhecimento de $(C C_g)^{\varepsilon}\varphi$ $(C C_g)^{\Diamond}\varphi$.

■

Corolário 5.4.1 (MOSES, 1986): No problema do ataque coordenado, qualquer protocolo que garanta que se uma divisão atacar, a outra divisão eventualmente atacará, é um protocolo em que nenhuma das divisões atacam. ■

prova: Seja (P_A,P_B) um *joint* protocolo que garanta que se uma das divisões atacar então ambas irão eventualmente atacar. Seja S e t_1 definidos como na prova da proposição (5.5.1). Seja φ = "O general A ou começou a atacar ou eventualmente irá atacar, e o general B ou começou a atacar ou eventualmente irá atacar". Pela descrição do problema, $(C C_g)^{\Diamond}\varphi$ é indeterminado em (r,t_1) para toda *run* $r \in S$. Devido as propriedades de (P_A,P_B), está claro que um general em ataque tem o conhecimento de $(C C_g)^{\Diamond}\varphi$ sob a interpretação de visão total. Logo, pelo teorema 5.4.2, o protocolo (P_A,P_B) garante que nenhum general irá atacar. ■

Teorema 5.5.1 (MOSES, DOLEV e HALPERN, 1985): Se existem n maridos infiéis em Mamajorca, no instante em que a rainha Henrietta I leu o documento, então as esposas traidas

atirarão em seus maridos na noite do n-ésimo dia. ■

prova: Se dá por indução no número n de maridos infiéis.

1. Para n = 1, já foi visto que o teorema é válido.

2. Suponha a validade do teorema para n = k. Deseja-se provar a valdade para n = k+1: assuma que existem k+1 maridos infiéis. Logo,suas esposas tem o conhecimento de k maridos infiéis e esperam ouvir tiros na k-ésima noite. Como esta noite passou em silêncio, elas concluem a existência de mais um marido infiel, que seria seu próprio marido. Assim, por indução, na (k+1)-ésima noite, serão ouvidos k+1 tiros. ■

Teorema 5.5.3 (MOSES, DOLEV e HALPERN, 1986): Usando um canal de comunicação síncrono, as esposas traidas que receberam a carta da rainha no dia significativo atirarão em seus maridos (n-1)d dias após o dia significativo, onde n é o número de maridos infiéis. Todas as outras esposas permanecerão na dúvida quanto a fidelidade de seus maridos.

prova: Suponha que uma esposa que tenha o conhecimento de k > 1 maridos infiéis não saiba inicialmente se existem k ou k+1 maridos infiéis. O que esta esposa tem conhecimento é que o dia significativo se encontra entre d-1 dias antes dela receber a carta e d-1 dias após ela receber a carta. Considerando k maridos infiéis, seu raciocínio seria: no mínimo uma das esposas traidas atiraria na noite ((k-1)d + 1) após o dia significativo, onde (k-1) representa o número de maridos infiéis conhecido pela esposa traida que atirou. Esta noite se encontra entre a noite ((k-2)d + 1 e a noite kd após o dia em que a esposa recebeu a carta. No caso de

haver k+1 maridos infiéis, o primeiro tiro será ouvido entre a noite ((k-1)d +2) ((k-1)d noites em silêncio é descoberta a existência de k-1 maridos infiéis e, acrescentando duas noites de silêncio, tem-se k-1 + 2 = k+1 maridos infiéis) e a noite kd + 1 após o dia em que a esposa recebeu a carta (kd+1 indica que a esposa recebeu a carta no dia significativo e desta forma, kd noites em silêncio é descoberto k maridos infiéis, com mais uma noite, tem-se a descoberta de k+1 maridos infiéis). Logo, se o tiro ocorre entre as noites ((k-1)d +2) e a noite kd após o dia em que a esposa recebeu a carta, esta, que tem o conhecimento de k maridos infiéis, ficará em dúvidas quanto a fidelidade de seu marido. ■

REFERÊNCIAS BIBLIOGRÁFICAS

[1] ANDERSON, A. e BELNAP, N. (1975), "Entailment, The Logic of Relevance and Necessity",Princenton University Press, Princenton.

[2] CASANOVA, M. A. ; GIORNO, F. C. e FURTADO, A. L. (1988), "Programação Lógica e a Linguagem PROLOG", Editora Edgar Bluncher Ltda.

[3] CHANG, C. L. e LEE, R. C. T. (1971), "Simbolic Logic and Mechanical Theorem Proving", Academic Press Inc.

[4] DA SILVA, J. C. P. (1990), "Logicas Não-Monotônicas na Formalização do Senso-Comum", Tese M. Sc., Engenharia de Sistema e Computação, COPPE/UFRJ.

[5] FAGIN, R. e HALPERN, J. (1988a), "Belief, Awareness, and Limited Reasoning", Artificial Intelligence, Vol.34, pp 39-76.

[6] FAGIN, R. e HALPERN, J. (1988b), "Reasoning About Knowledge and Probability Preliminary Report", Procedings of the second conference of Theoretical Aspects of Reasoning About Knowledge (TARK 1988), pp 277-293.

[7] FAGIN, R. e HALPERN, J. (1988c), "Uncertainty, Belief and Probability",IBM Research Report, RJ 6191(60901).

[8] FAGIN, R.; HALPERN, J. e VARDI, M. (1984), "A Model-Theoretic Analysis of Knowledge: Preliminary Report", IBM Research Report, RJ 4373(47631).

[9] FAGIN, R.; HALPERN, J. e VARDI, M. (1988), "What can Machines Know ? On The Properties of Knowledge in Distributed Systems", IBM Research Report, RJ 6250 (61647).

[10] FAGIN, R.;HALPERN, J. e VARDI, M. (1990), "A Nonstandard Approach to the Logical Omniscience Problem",

Proceedings of the Third Conference of Theoretical Aspects of Reasoning About Knowledge (Tark 1990),pp 41-55 - Pacific Grove, CA.

[11] FAGIN, R. e VARDI, M. (1986), "Knowledge and Implicit Knowledge in a Distributed Envirorirment :Preliminary Report", IBM Research Report, RJ 4990 (52167).

[12] HALPERN, J (1987), "Using Reasoning About Knowledge to Analize Distributed Systems", IBM Research Report, RJ 5522 (56421).

[13] HALPERN, J. (1989), "The Relationship Between Knowledge, Belief and Certainty", IBM Research Report, RJ 6765(64833).

[14] HALPERN, J. e MOSES, Y. (1984a), "Knowledge and Common Knowledge in a Distributed Environment", IBM Research Report, RJ 4421 (47909). [15] HALPERN, J. e MOSES, Y. (1984b), "Towards a Theory of Knowledge and Ignorance : Preliminary Report", IBM Research Report, RJ 4448 (48136).

[16] HALPERN, J. e MOSES, Y. (1985), "A Guide to the Modal Logic of Knowledge and Belief : Preliminary Draft", RJ 4753 (50521).

[17] HALPERN, J. e ZUCK. D. (1987), "A Little Knowledge Goes a Long Way : Simple Knowledge-based Derivations and Carrecteness Proofs for a Family of Protocols", IBM Research Report, RJ 5857 (58908).

[18] HINTIKKA, J. (1962), " Knowledge and Belief", Cornell University Press, Cornell.

[19] HUGHES, G. E. e CRESWELL, M. J. (1968), "An Introduction to Modal Logic", Methuen and Co LTD.

[20] LEVESQUE, H. (1984), "A Logic of Implicit and Explicit Belief", Proceedings of National Conference on Artificial Intelligence (AAAI-84), pp 198-202.

[21] LEVESQUE, H.(1990), "All I Know : A Study in Autoepistemic Logic", Artificial Intelligence, Vol.42, pp 263-309.

[22] McARTHUR, G. (1988), "Reasoning About Knowledge and Belief: a Survey", Computational Inteligence, Vol4, pp 223-243.

[23] McCARTHY, J. (1980), "Circumscription : A Form of Non-Monotonic Reasoning", Artificial Intelligence, Vol.13, pp 27-39.

[24] McDERMOTT, D. e DOYLE, J. (1980), "Non-Monotonic Logic I", Artificial Intelligence, Vol.13, pp 41-72.

[25] MOORE, R. C. (1985), "Semantical Considerations on Nom-Monotonic Logic", Artificial Intelligence, Vol.25, pp 75-94.

[26] MOSES, Y.; DOLEV, D. e HALPERN, J. ?. (1985), "Cheating Husbands and Other Stories : A Case Study of Knowledge, Action, and Comunication", IBM Research Report, RJ 4756 (50524).

[27] MOSES, Y. O. (1986), "Knowledge in a Distributed Environment",Phd Thesis, Department of Computer Science, Stanford University.

[28] NGUYEN, V. e PERRY K. J. (1986), "Do You Really Know what Knowledge Is ?", IBM Research Report, RC 11830 (Log # 51378).

[29] RAO, A. S. e FOO, N. Y. (1987a),"Envolving Knowledge and Logical Omniscience", IBM Research Report, RC 13155 (# 59692).

[30] RAO, A. S. e FOO, N. Y. (1987b), "Envolving Knowledge and Autoepistemic Reasoning", IBM Research Report, RC 13156 (#59695).

[31] REITER, R. (1980), "A logic for Default Reasoning",

Artificial Intelligence, Vol.13, pp 81-132.

[32] REITER, R (1988), "Non-Monotonic Reasoning", Exploring Artificial Intelligence : Survey Talks from the National Conference on A.I.", pp 439-481.

[33] SHOHAM, Y. e MOSES, Y., "Belief as Defeasible Knowledge", Proceedings of the 1st Conference on Principles of Knowledge Representation and Reasoning, pp 1168-1172, Toronto-Canadá, 1989.

[34] ZADROZNY, W. (1986), "Explicit and Implicit Beliefs, Preliminary Version", IBM Research Report, RC 11786 (# 52921).