

Thiago Ângelo Gelaim

**MODELO DE AGENTES E-BDI INTEGRANDO CONFIANÇA  
BASEADO EM SISTEMAS MULTI-CONTEXTO**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Ciência da Computação.

Orientador: Prof. Ricardo Azambuja Silveira, Dr.

Coorientadora : Prof<sup>a</sup>. Jerusa Marchi, Dr<sup>a</sup>.

Florianópolis

2016

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Gelaim, Thiago Ângelo  
Modelo de Agentes E-BDI Integrando Confiança Baseado em  
Sistemas Multi-Contexto / Thiago Ângelo Gelaim ;  
orientador, Ricardo Azambuja Silveira ; coorientadora,  
Jerusa Marchi. - Florianópolis, SC, 2016.  
93 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro Tecnológico. Programa de Pós-Graduação em  
Ciência da Computação.

Inclui referências

1. Ciência da Computação. 2. Agente. 3. Confiança. 4.  
Sistema Multi-Contexto. 5. Emoção. I. Silveira, Ricardo  
Azambuja. II. Marchi, Jerusa. III. Universidade Federal de  
Santa Catarina. Programa de Pós-Graduação em Ciência da  
Computação. IV. Título.

Thiago Ângelo Gelaim

**MODELO DE AGENTES E-BDI INTEGRANDO CONFIANÇA  
BASEADO EM SISTEMAS MULTI-CONTEXTO**

Esta Dissertação foi julgada aprovada para a obtenção do Título de “Mestre em Ciência da Computação”, e aprovado em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Florianópolis, 05 de fevereiro 2016.

---

Prof<sup>a</sup>. Carina Friedrich Dorneles, Dr<sup>a</sup>  
Coordenadora do Programa

---

Prof<sup>a</sup>. Jerusa Marchi, Dr<sup>a</sup>.  
Coorientadora

**Banca Examinadora:**

---

Prof. Ricardo Azambuja Silveira, Dr.  
Universidade Federal de Santa Catarina  
Orientador

---

Prof<sup>a</sup>. Diana Francisca Adamatti, Dr<sup>a</sup>  
Universidade Federal do Rio Grande

---

Prof. Luiz Antônio Moro Palazzo, Dr.  
Universidade Católica de Pelotas

---

Prof. Elder Rizzon Santos, Dr.  
Universidade Federal de Santa Catarina



## AGRADECIMENTOS

Agradeço à minha companheira, Naiane Cristina Salvi, fundamental nesta conquista.

À minha família, em especial aos meus pais, pelos ensinamentos que levo para a vida.

Ao meu orientador, Ricardo Azambuja Silveira, pelas inúmeras contribuições no meu desenvolvimento acadêmico.

À minha coorientadora, Jerusa Marchi, pelas preciosas discussões teóricas.

Aos professores das bancas de qualificação e de defesa pelas contribuições na pesquisa.

Aos meus amigos e colegas do grupo de pesquisa IATE pelos conhecimentos compartilhados.

Aos meus amigos recentes e antigos, de perto e de longe.

À CAPES pelo apoio financeiro.

À todas as pessoas e entidades que direta ou indiretamente colaboraram na minha formação.



## RESUMO

Em sistemas multiagente, seus integrantes precisam interagir a fim de realizar seus objetivos. Existem situações em que entidades mal-intencionadas fazem parte do ambiente. Desta forma, para decidir com quem interagir é preciso considerar o comportamento dos candidatos. Neste contexto, mecanismos de confiança são utilizados como forma de proteção contra entidades fraudulentas. Na literatura são encontrados diversos modelos de confiança, em muitos casos considerando mais o aspecto funcional do modelo do que a sua influência no ciclo de raciocínio do agente. De acordo com a perspectiva do agente, um modelo de confiança é como uma “caixa-preta”. O objetivo deste trabalho é propor um modelo de agente em que a confiança faça parte de seu raciocínio. O modelo proposto, inspirado em BDI, é definido a partir de um sistema multi-contexto e faz uso de um modelo computacional de emoções, visando vincular a subjetividade das experiências do agente com a definição de confiança.

**Palavras-chave:** Agente. Confiança. Emoção. Sistema Multi-Contexto.





## ABSTRACT

In multiagent systems, its members need to interact in order to fulfill its goals. There are situations where malicious entities are inhabiting the environment. Thus, to decide with whom to interact it is necessary to consider the behavior of the candidates. In this context, trust mechanisms are used as a protection against fraudulent entities. The literature presents various trust models, in many cases considering more functional aspect of the model than its influence on the agent's reasoning cycle. According to the agent's perspective, a trust model is like a "black box". The goal of this work is to propose an agent model where trust is part of its reasoning. The proposed model, inspired by BDI, is defined as a multi-context system and makes use of a computational model of emotions aiming to link the subjectivity of the agent's experiences with the definition of trust.

**Keywords:** Agent. Trust. Emotion. Multi-Context System.



## LISTA DE FIGURAS

Figura 1	Estágios e camadas de confiança. ....	29
Figura 2	Arquitetura do modelo Repage. ....	32
Figura 3	Modelo BDI com Emoções e Confiança. ....	55
Figura 4	Modelo de agente proposto. Cada aresta (ou grupo de arestas) representa uma regra de ponte; a numeração refere-se a regra correspondente listada na seção 4.8. ....	58
Figura 5	Dinâmica emocional do modelo WASABI. ....	61
Figura 6	Emoções primárias no espaço PAD. ....	63
Figura 7	Estrutura taxonômica presente no contexto de confiança. ....	81



## **LISTA DE TABELAS**

Tabela 1	Tipos de crenças utilizadas para confiança. ....	31
Tabela 2	Tipos de crenças para delegação fraca. ....	31
Tabela 3	Valores para estímulos emocionais. ....	79



## LISTA DE ABREVIATURAS E SIGLAS

BDI	<i>Belief-Desire-Intention</i> . . . . .	19
PRS	<i>Procedural Reasoning System</i> . . . . .	25
E-BDI	<i>Emotional-Belief-Desire-Intention</i> . . . . .	25
SMA	Sistema Multiagente . . . . .	26
SMC	Sistema Multi-Contexto . . . . .	27
IA	<i>Inteligência Artificial</i> . . . . .	40
WASABI	<i>[W]ASABI [A]ffect [S]imulation for [A]gents with [B]elievable [I]nteractivity</i> . . . . .	40
OCC	Ortony, Clore e Collins . . . . .	40
PAD	<i>Pleasure-Arousal-Dominance</i> . . . . .	40
LDPO	Lógica dinâmica de Primeira Ordem . . . . .	51
fbfs	fórmulas bem formadas . . . . .	52





## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	19
1.1	OBJETIVOS	20
<b>1.1.1</b>	<b>Objetivo Geral</b>	20
<b>1.1.2</b>	<b>Objetivos Específicos</b>	20
1.2	METODOLOGIA	21
1.3	ORGANIZAÇÃO DO TRABALHO	22
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	23
2.1	AGENTES	23
<b>2.1.1</b>	<b>Agentes BDI</b>	24
<b>2.1.2</b>	<b>Agentes E-BDI</b>	25
<b>2.1.3</b>	<b>Sistemas Multiagente</b>	26
2.2	AMBIENTE	26
2.3	SISTEMAS MULTI-CONTEXTO	27
2.4	CONFIANÇA	28
<b>2.4.1</b>	<b>Modelos de Confiança e Reputação</b>	30
2.4.1.1	Castelfranchi e Falcone	30
2.4.1.2	Repage	32
2.4.1.3	ReGreT	33
2.4.1.4	FIRE	38
2.5	EMOÇÕES	39
<b>2.5.1</b>	<b>WASABI</b>	41
<b>2.5.2</b>	<b>Steunebrink e TrustE</b>	42
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	45
3.1	GRADED BDI	45
3.2	BDI + REPAGE	48
3.3	ADAPTRUST	50
3.4	OUTROS TRABALHOS	54
3.5	ANÁLISE DOS TRABALHOS RELACIONADOS	55
<b>4</b>	<b>MODELO PROPOSTO</b>	57
4.1	CONTEXTO DE COMUNICAÇÃO (CC)	58
4.2	CONTEXTO DE EMOÇÕES (EC)	59
<b>4.2.1</b>	<b>Atualização do Estado Emocional</b>	60
4.2.1.1	Dinâmica Emocional	60
4.2.1.2	Mapeamento para o Espaço PAD	62
4.3	CONTEXTO DE CRENÇAS (BC)	63
<b>4.3.1</b>	<b>A Linguagem de Crenças</b>	64
4.4	CONTEXTO DE CONFIANÇA (TC)	66

<b>4.4.1</b>	<b>Adicionando o Modelo ReGreT</b> .....	68
4.5	CONTEXTO DE DESEJOS (DC) .....	69
<b>4.5.1</b>	<b>A Linguagem de Desejos</b> .....	69
4.6	CONTEXTO DE INTENÇÕES (IC) .....	70
<b>4.6.1</b>	<b>A Linguagem de Intenções</b> .....	70
4.7	CONTEXTO DE PLANEJAMENTO (PC) .....	71
4.8	REGRAS DE PONTE .....	72
<b>5</b>	<b>DINÂMICA DO MODELO</b> .....	75
5.1	PROCURANDO POR UM MECÂNICO .....	75
5.2	TUTORIA .....	78
<b>5.2.1</b>	<b>Dinâmica do Agente</b> .....	81
5.2.1.1	Criação da Crença de Confiança .....	83
<b>6</b>	<b>CONCLUSÃO</b> .....	85
6.1	TRABALHOS FUTUROS .....	86
	<b>REFERÊNCIAS</b> .....	87

## 1 INTRODUÇÃO

Indivíduos pertencentes a uma sociedade, seja ela real ou virtual, geralmente interagem para realizar seus objetivos. Estas interações visam solicitar serviços ou obter recursos de outros indivíduos. Um ambiente pode conter uma ou mais entidades capazes, ou que se dizem capazes, de cumprir determinada solicitação. Portanto, decidir interagir com alguém implica em delegar responsabilidades e confiar neste alguém. Esta decisão de confiar ou não pode ser baseada tanto no conhecimento a respeito das capacidades do agente avaliado, quanto nas opiniões obtidas de outros agentes sobre estas capacidades.

Em uma sociedade virtual, os agentes de um sistema multiagente podem ser projetados por diferentes indivíduos, e com isso, podem possuir objetivos distintos (WOOLDRIDGE, 2009). Além disto, entidades com intenções fraudulentas podem ser instanciadas. Vários modelos para confiança e reputação foram desenvolvidos visando avaliar comportamentos e melhorar a interação entre os agentes (SABATER-MIR; PAOLUCCI, 2007).

Como apontado por Sabater-Mir e Sierra (2012, p. 417), os esforços para projetar melhores modelos de confiança e reputação estão sendo direcionados para integrar estes componentes a outros processos executados pelo agente. Isto é, a confiança está deixando de ser considerada uma “caixa-preta” e começando a ser integrada ao raciocínio do agente.

Em um agente *belief-desire-intention* (BDI), as crenças (*beliefs*) representam o que o agente acredita sobre o mundo, os desejos (*desires*) representam o seu estado motivacional; e as intenções (*intentions*) constituem o seu componente deliberativo (RAO; GEORGEFF et al., 1995). Integrar confiança em BDI tem a vantagem da clara definição de crenças e desejos do agente, permitindo que o papel desempenhado pela confiança seja feito de forma explícita (KOSTER; SCHORLEMMER; SABATER-MIR, 2012). Por exemplo, para realizar o desejo de *aprenderAutômatos*, o agente precisa acreditar na existência de uma entidade confiável, capaz de ensinar autômatos em conformidade com propriedades estabelecidas, tais como disponibilidade e método de ensino.

Castelfranchi e Falcone (2010, p.133) apontam que existe um lado afetivo na confiança. Este “aspecto afetivo pode desempenhar um papel modificando o processo de crença [...]”.<sup>1</sup> E também, pode transformar a confiança como uma “reação emocional não baseada em crenças” (CASTELFRANCHI; FALCONE, 2010) por meio de marcadores somáticos (DAMÁSIO, 1996). De acordo com a hipótese de marcadores somáticos, proces-

---

<sup>1</sup>The affective aspect of trust can play a role by modifying the belief process [...]

tos emocionais ajudam na tomada de decisão dando destaque a algumas opções (DAMÁSIO, 1996).

Como o lado afetivo é pouco explorado na definição de confiança em agentes, esta pesquisa integra confiança e emoções em um modelo de agentes BDI, visando aprimorar o seu mecanismo de deliberação. A abordagem de sistemas multi-contexto é utilizada para integrar confiança e emoções com as crenças, desejos e intenções do agente. Um sistema multi-contexto permite que diferentes componentes formais sejam definidos e se relacionem (GIUNCHIGLIA; SERAFINI, 1994).

## 1.1 OBJETIVOS

Embora exista relação entre emoções e confiança, na literatura de agentes, a integração não atingiu sua maturidade. Com isto, esta pesquisa objetiva contribuir nesta integração.

### 1.1.1 Objetivo Geral

Propor um modelo de agentes, baseado na teoria de crenças, desejos e intenções (BDI) que integre aspectos emocionais no raciocínio sobre confiança, seguindo a abordagem de sistemas multi-contexto.

### 1.1.2 Objetivos Específicos

Visando a concretização do objetivo geral, os seguintes objetivos específicos são definidos:

- Analisar modelos cognitivos e numéricos de confiança;
- Analisar modelos computacionais de emoções;
- Definir um modelo conceitual BDI, seus contextos e relacionamentos;
- Desenvolver um contexto para reger as informações sobre confiança e integrá-lo aos componentes do agente;
- Desenvolver um contexto para reger as informações sobre emoções e integrá-lo as demais partes do agente;
- Descrever a dinâmica do modelo proposto por meio de cenários de teste.

## 1.2 METODOLOGIA

Esta é uma pesquisa construtiva (*design science*), seguindo o modelo proposto por Takeda, Veerkamp e Yoshikawa (1990). As cinco fases desta abordagem são apresentadas a seguir:

### 1. Compreensão do Problema

Os esforços da fase de compreensão do problema estão na delimitação do escopo do problema. O resultado desta etapa é a proposta, formal ou informal, da pesquisa. Assim, os temas estudados são:

- (a) Agentes e Sistemas Multiagente: revisão bibliográfica da área; teorias, arquiteturas e linguagens de agentes;
- (b) Confiança e reputação: perspectivas de diversas áreas; modelos computacionais de confiança; modelos/arquiteturas de agentes integrando confiança;
- (c) Emoções: visão geral sobre o tema; modelos computacionais de emoções; modelos/arquiteturas de agentes integrando emoções;
- (d) Confiança e emoções: relação entre confiança e emoções.

### 2. Sugestão

A sugestão é a fase seguinte à compreensão do problema, nela são geradas sugestões para a solução do problema. Duas soluções foram levantadas: (i) um módulo genérico para confiança e reputação com emoções para ser integrado em agentes BDI ou E-BDI; e (ii) um modelo de agentes.

### 3. Desenvolvimento

Esta fase é destinada ao desenvolvimento e implementação da solução de acordo com as sugestões. Como levantado anteriormente, foram sugeridas duas soluções:

- Um módulo: esta abordagem durante o desenvolvimento se tornou inviável, o principal motivo foi garantir a interoperabilidade entre o módulo e os demais componentes do agente;
- Um modelo de agentes: sugestão desenvolvida de acordo com sistemas multi-contexto. A escolha por esta abordagem foi motivada pela possibilidade de integrar diversos componentes utilizando regras de ponte. Estas regras permitem a troca de informações entre os componentes.

#### 4. Avaliação

É nesta fase que o modelo proveniente da fase anterior é testado. Na literatura existem plataformas específicas para testes de modelo de confiança e reputação, como por exemplo, a plataforma ART (*Agent Reputation and Trust testbed*) (FULLAM et al., 2005) e a plataforma TREET (*TREET: The trust and reputation experimentation and evaluation testbed*) (KERR; COHEN, 2010). Porém, nesta pesquisa, optou-se por desenvolver cenários priorizando descrever o comportamento do agente.

#### 5. Conclusão

O objetivo desta fase é consolidar o conhecimento adquirido nas fases anteriores de forma mais concreta. Nela, a integração de confiança e emoções em um agente BDI utilizando sistema multi-contexto é analisada.

### 1.3 ORGANIZAÇÃO DO TRABALHO

No capítulo 2 são apresentados os principais conceitos relacionados a pesquisa. No capítulo 3 são abordados trabalhos que se assemelham com esta pesquisa. O modelo de agente proposto é descrito no capítulo 4. Avaliações são feitas no capítulo 5. Por fim, as considerações finais e trabalhos futuros são apresentadas no capítulo 6.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os principais conceitos que norteiam a pesquisa. A entidade *agente* é descrita na seção 2.1, sendo enfatizado o modelo BDI. *Sistema multiagente* e *ambiente* são apresentados nas seções 2.1.3 e 2.2, respectivamente, visando contextualizar o modelo proposto. *Sistemas multi-contexto* são definidos na seção 2.3 como uma ferramenta para corporificar a proposta. Os temas *confiança* e *emoções* são descritos, respectivamente, nas seções 2.4 e 2.5 objetivando sua integração em agentes BDI.

### 2.1 AGENTES

Agente é um sistema computacional situado em um ambiente onde ele é capaz de realizar ações autônomas, visando cumprir os seus objetivos de projeto (WOOLDRIDGE, 2002). O agente é autônomo quando consegue perceber e atuar sobre o ambiente, ao longo do tempo, de acordo com a sua agenda e de forma a afetar o que ele irá perceber no futuro (FRANKLIN; GRAESSER, 1997).

Um agente racional é aquele que escolhe executar ações de acordo com seus interesses, considerando as crenças que ele tem sobre o mundo (WOOLDRIDGE, 2000). Por exemplo, se o seu objetivo é ficar seco, e ele acredita que está chovendo, é racional utilizar um guarda-chuva quando estiver saindo de casa (WOOLDRIDGE, 2000). O agente inteligente é capaz de executar ações reativas, apresenta comportamento pró-ativo e habilidades sociais (WOOLDRIDGE, 2002).

Wooldridge e Jennings (1995) organizam o projeto e a construção de agentes em três áreas: teorias, arquiteturas, e linguagens. Teorias são basicamente especificações, elas buscam definir o que o agente é, as propriedades que deve apresentar, como raciocinar sobre estas, etc. Arquiteturas são modelos de engenharia de software para agentes, de acordo com elas, o foco está em satisfazer as propriedades especificadas pelas teorias. Linguagens permitem a programação e execução de agentes, elas podem incluir princípios e propriedades estabelecidos pelas teorias (WOOLDRIDGE; JENNINGS, 1995).

### 2.1.1 Agentes BDI

O modelo *Belief-Desire-Intention* (BDI) é baseado na teoria de raciocínio prático proposta por Bratman (1987). O raciocínio prático é o raciocínio direcionado a ações e envolve dois processos principais: decidir quais objetivos alcançar (deliberação); e decidir como alcançar (raciocínio meio-e-fim) (WOOLDRIDGE, 2000).

Bratman (1987) argumenta que intenções possuem grande importância no raciocínio prático. Wooldridge (2000) resume os papéis delas em:

1. Intenções direcionam o raciocínio meio-e-fim: dado que uma intenção está formada, o agente tentará realizá-la. Para isto, ele precisa decidir *como* realizar, e se um curso de ações falhar, ele pode tentar outros cursos.
2. Intenções persistem: tipicamente intenções persistirão até que uma destas três situações ocorra: (i) o agente acredite que ela foi realizada com sucesso; (ii) o agente acredite que ela não pode ser realizada; e (iii) a razão que formou a intenção não existe mais.
3. Intenções restringem deliberações futuras: o agente não irá escolher opções que são inconsistentes com as intenções atuais.
4. Intenções influenciam as crenças sobre as quais raciocínios práticos futuros são baseados: uma vez adotada uma intenção, planos futuros podem considerar que o agente irá realizar a intenção.

Cohen e Levesque (1990) criaram uma teoria de intenções em que elas são resultados dos desejos e crenças do agente, ou seja, apenas crenças e desejos são atitudes mentais primitivas. Rao, Georgeff et al. (RAO; GEORGEFF, 1991; RAO; GEORGEFF et al., 1995) apresentam uma formalização para agentes BDI mais próxima ao trabalho de Bratman (MEYER et al., 2014), onde crenças, desejos e intenções são atitudes mentais primitivas.

A forma como estes três componentes são combinados descreve o comportamento do agente BDI. Rao, Georgeff et al. (1995) apresentam os termos de realismo fraco (*weak realism*), realismo (*realism*) e realismo forte (*strong realism*). De acordo com o realismo fraco, o agente não deseja uma proposição se ele acredita na sua negação e não tem a intenção de uma proposição se ele deseja ou acredita na negação. Segundo o realismo, as proposições que o agente acredita serem possíveis são desejadas e intencionadas. Para o realismo forte, se o agente não acredita em uma proposição, ele não terá o desejo e nem a intenção disto, se o agente deseja uma proposição,



ele vai acreditar que a proposição é uma opção (PARSONS; SIERRA; JENNINGS, 1998).

Um dos principais exemplos de arquitetura BDI é a PRS (*Procedural Reasoning System*) (GEORGEFF; LANSKY, 1986). Ela foi uma das primeiras arquiteturas BDI, e serviu como base para outras como UM-PRS, C-PRS, AgentSpeak(L), dMARS e JAM (WOBCKE, 2007). Segundo a arquitetura PRS, o agente tem representação explícita de estruturas de dados que correspondem, levemente, à crenças, desejos e intenções, além de uma biblioteca pré-compilada de planos (WOOLDRIDGE, 2009). O algoritmo 1 apresenta o ciclo básico de controle em um agente BDI segundo Wooldridge (2000).

---

**Algoritmo 1:** Ciclo básico de raciocínio de um agente BDI

---

```

1 enquanto verdade faça
2   observe o mundo
3   atualize o modelo interno do mundo
4   delibere sobre qual intenção alcançar
5   utilize raciocínio meio-e-fim para obter um plano para a
   intenção
6   execute o plano
7 fim

```

---

### 2.1.2 Agentes E-BDI

A adição de emoções em agentes BDI vem sendo explorada pelo menos desde 2004 no trabalho de Hernandez et al. (2004). A clara distinção dos componentes do agente facilita a modelagem e implementação de emoções (HERNANDEZ et al., 2004). A influência que elas exercem em agentes BDI pode ser explorada tanto na deliberação quanto no raciocínio meio-e-fim (PEREIRA et al., 2005; JIANG; VIDAL; HUHNS, 2007; CÔRTEZ, 2013; PUICĂ; FLOREA, 2013).

Por exemplo, um agente E-BDI situado em um ambiente, pode utilizar suas emoções para focar seus sensores em um ponto específico e, com isto, gerar crenças sobre a parte específica do ambiente. A revisão de crenças também pode sofrer influência emocional. Por exemplo, o agente pode priorizar um subconjunto de crenças que o deixou mais “feliz”. O mesmo raciocínio pode ser aplicado para priorizar desejos e selecionar intenções (WOOLDRIDGE, 2000; JIANG; VIDAL; HUHNS, 2007; SIGNORETTI, 2012; PUICĂ; FLOREA, 2013; CÔRTEZ, 2013; LEJMI-RIAAHI; KEBAIR; SAID,

2014).

O modelo proposto estende o modelo BDI analisando a influência que emoções tem na definição da confiança. Emoções só influenciam nas crenças relacionadas aos julgamentos do agente. Porém, elas são influenciadas pelas percepções, crenças e intenções do agente.

### 2.1.3 Sistemas Multiagente

Um Sistema MultiAgente (SMA) pode ser descrito como um sistema composto por uma coleção de agentes. Os agentes de um SMA podem interagir por meio de cooperação, coordenação e negociação. Cooperar é trabalhar em conjunto para cumprir um objetivo comum. Na coordenação a resolução do problema é organizada de forma a evitar atividades prejudiciais e explorar atividades benéficas. Negociar é chegar a um acordo aceitável para todas partes envolvidas (WOOLDRIDGE, 2009; JENNINGS; SYCARA; WOOLDRIDGE, 1998). Em um SMA cada agente possui informações e/ou capacidades limitadas para resolver o problema, não existe um sistema de controle global, os dados são descentralizados e a computação é assíncrona (JENNINGS; SYCARA; WOOLDRIDGE, 1998).

Um sistema multiagente aberto, permite que, a qualquer momento, agentes possam entrar e sair. Além disto, agentes em um SMA aberto podem ser inseridos por diferentes partes interessadas, com diferentes metas e objetivos. Assim, pode ser assumido que em um SMA aberto: (i) as ações dos agentes são voltadas ao próprio interesse, e com isto, pouco confiáveis; (ii) nenhum agente pode conhecer tudo sobre o ambiente; (iii) os agentes não são controlados por uma autoridade central (HUYNH; JENNINGS; SHADBOLT, 2006).

## 2.2 AMBIENTE

Como apresentado na seção 2.1, agentes estão situados em um ambiente. Russell, Norvig e Souza (2004) organizam os ambientes nos seguintes tipos:

- Completamente observável ou parcialmente observável: um ambiente é completamente observável se os sensores do agente podem obter de forma precisa, completa e atualizada as informações do estado do ambiente. Caso existam ruídos ou imprecisões, o ambiente é parcialmente observável;

- Determinístico ou estocástico: um ambiente é determinístico se o próximo estado do ambiente é completamente determinado pelo estado atual e pela ação executada pelo agente. Caso contrário, o ambiente é estocástico;
- Episódico ou sequencial: em um ambiente episódico, a experiência do agente é organizada em episódios atômicos. O episódio seguinte não pode depender das ações executadas em episódios anteriores. Em ambientes sequenciais, a decisão atual pode afetar decisões futuras;
- Estático ou dinâmico: um ambiente estático permanece inalterado até que o agente execute alguma ação sobre ele. Um ambiente dinâmico pode mudar enquanto o agente está deliberando sua próxima ação;
- Discreto ou contínuo: a separação entre discreto e contínuo pode ser feita a partir dos estados do ambiente, das percepções e ações do agente e do tempo. Por exemplo, o ambiente de um jogo de xadrez tem um número discreto de estados, por sua vez, dirigir um táxi é formado por um número contínuo de estados.

O conhecimento dos tipos de ambiente é importante para o desenvolvimento do modelo proposto, pois as percepções podem influenciar tanto no estado emocional, quanto na confiança. Por ser baseado em BDI e de acordo com esta categorização de ambientes, é possível que o agente proposto habite na pior situação, ou seja, em um ambiente parcialmente observável, estocástico, sequencial, dinâmico e contínuo.

## 2.3 SISTEMAS MULTI-CONTEXTO

Um sistema multi-contexto (SMC) é um *framework* que possibilita que diferentes componentes teóricos (contextos) sejam definidos e inter-relacionados (GIUNCHIGLIA; SERAFINI, 1994). Na modelagem de agentes como SMC, um contexto é construído com uma única lógica e suas teorias, a fim de representar uma unidade principal da arquitetura. Regras de ponte são regras de inferências que relacionam fórmulas de diferentes contextos. (PARSONS; SIERRA; JENNINGS, 1998; NORIEGA; SIERRA, 1997; CASALI et al., 2008; CASALI; GODO; SIERRA, 2005).

**Definição 2.1** (Agente como Sistema Multi-Contexto). Seja  $I$  um conjunto finito de índices. Um agente é um par ordenado

$$\langle \{C_i\}_{i \in I}, \Delta_{rp} \rangle,$$

onde  $\{C_i\}_{i \in I}$  é um conjunto de contextos e  $\Delta_{rp}$  é um conjunto de regras de ponte. Cada contexto  $C_i$  é uma tupla

$$\langle L_i, A_i, \Delta_i \rangle,$$

onde  $L_i$  é uma linguagem,  $A_i$  é o conjunto de axiomas e  $\Delta_i$  são as regras de inferência.

Por exemplo, sejam  $C_1$ ,  $C_2$  e  $C_3$  contextos,  $\psi$  e  $\varphi$  fórmulas dedutíveis de  $C_1$  e  $C_2$  respectivamente. Com a regra de ponte 2.1, a fórmula  $\theta$  é adicionada ao contexto  $C_3$ .

$$\frac{C_1 : \psi, C_2 : \varphi}{C_3 : \theta} \quad (2.1)$$

Segundo Sabater et al. (2000), a utilização de sistemas multi-contexto em agentes tem vantagens do ponto de vista da engenharia de software e da modelagem lógica. De acordo com a primeira, SMC apoiam a decomposição modular e encapsulamento, ou seja, cada componente da arquitetura — seja ele funcional ou uma estrutura de dados, pode ser representado como um contexto que pode ser reutilizado por outro contexto. Para a segunda, as vantagens estão na escalabilidade e eficiência geradas pela divisão em contextos e na possibilidade de utilização de diferentes lógicas, o que permite o aumento do poder de representação e a simplificação conceitual de agentes.

Com base nessas vantagens e nos trabalhos encontrados na área, a integração de confiança e emoções por meio de SMC, de acordo com esta pesquisa, torna-se adequada. No capítulo 4 o modelo proposto é apresentado segundo esta abordagem.

## 2.4 CONFIANÇA

Em sociedades reais e virtuais, o conceito de confiança é um fator importante para execução de ações e em interações. Ela é uma característica altamente problemática, porém recorrente em relacionamentos sociais (MISZTAL, 2013). Confiança é tema de estudos em diversas áreas como por exemplo na psicologia, na sociologia, na filosofia, na economia e na computação.

Confiança, segundo Rousseau et al. (1998), é “um estado psicológico compreendendo a intenção de aceitar vulnerabilidade, baseada em expectativas positivas das intenções ou comportamento de outrem”. De acordo com Gambetta et al. (2000)

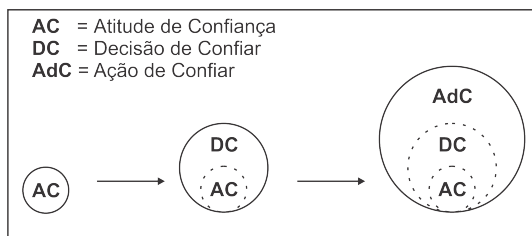
Confiança (ou simetricamente, desconfiança) é um ní-

vel particular da probabilidade com a qual um agente avalia outro agente ou grupo de agentes que irá/irão executar uma ação particular, tanto *antes* dele poder monitorar tal ação (ou independentemente da sua capacidade de monitorar) e em um contexto no qual isto afeta a sua própria ação.

Castelfranchi e Falcone (2010) analisam confiança a partir de camadas. Estas estão representadas na figura 1, e significam:

1. No seu sentido básico apenas uma atitude mental (cognitiva e afetiva) e disposição para o confiado;
2. Sob uma perspectiva mais incrementada, uma decisão e intenção baseada na disposição;
3. Então no ato de confiar no comportamento esperado do confiado;
4. E por fim a interação e relação social com o confiado.

Figura 1 – Estágios e camadas de confiança.



Fonte: Adaptado de Castelfranchi e Falcone (2010).

Enquanto confiança é algo próprio do agente, a reputação é uma “medida coletiva de confiabilidade baseada nas referências ou avaliações dos membros em uma comunidade” (JØSANG; ISMAIL; BOYD, 2007). Mecanismos de confiança e reputação são utilizadas por agentes como forma de proteção contra entidades fraudulentas. Em SMA, existem três abordagens complementares (CATADAU, 2010):

1. Segurança: assegura propriedades estruturais básicas, como por exemplo, autenticidade, integridade de mensagens, privacidade, identidade dos agentes. Esta segurança pode ser obtida por meios como criptografia e assinatura digital;

2. Institucional: abordagem em um nível de abstração mais alto que o nível de segurança. Nesta existe uma entidade central que observa, controla ou força as ações dos agentes. Comportamentos não desejáveis podem ser punidos;
3. Social: os próprios agentes são capazes de avaliar e punir comportamentos indesejados. Essa abordagem abriga os modelos de confiança e reputação.

De acordo com esta categorização, a abordagem social é explorada nesta pesquisa. Na literatura são encontrados vários modelos de confiança e reputação, Sabater e Sierra (2005) e Pinyol e Sabater-Mir (2013) fazem um levantamento apresentando as características principais destes. Por exemplo, um modelo pode ser cognitivo ou numérico; pode considerar ou não a presença de agentes trapaceiros no ambiente; as fontes de informação utilizadas podem ser obtidas com (i) experiências diretas (interações ou observações); (ii) informação de testemunha; (iii) informação sociológica (relações sociais dos agentes); (iv) preconceito (propriedades atribuídas a partir dos grupos que o agente faz parte).

## **2.4.1 Modelos de Confiança e Reputação**

Visando realizar o objetivo de analisar modelos cognitivos e numéricos de confiança. Nesta seção é feita uma breve descrição sobre alguns destes modelos. São apresentados os modelos cognitivos de Castelfranchi e Falcone e o modelo Repage. Eles foram escolhidos por já terem sido integrados em agentes BDI (HÜBNER et al., 2009; PINYOL; SABATER-MIR, 2009). Também são apresentados os modelos numéricos ReGreT e FIRE. O primeiro foi escolhido por adicionar a influência de emoções no trabalho de Bitencourt, Silveira e Marchi (2013), e o segundo por sofrer forte influência do primeiro. Além disso, a forma como a interação direta é abordada é igual nos dois modelos. Sendo que este é o principal ponto explorado no modelo proposto.

### **2.4.1.1 Castelfranchi e Falcone**

Castelfranchi e Falcone (2001) propõem um modelo cognitivo de confiança. Segundo os autores, a confiança é um estado mental composto por crenças e objetivos; “uma atitude complexa de um agente  $x$  para com um agente  $y$  sobre o comportamento/ação  $a$  relevante para o resultado (objetivo)

$g$ ". Três denotações são discutidas para a palavra "confiança":

- Confiança principal (*core trust*): avaliação do agente  $y$  antes de depender dele;
- Dependência (*reliance*): a confiança principal mais a decisão de depender do agente  $y$ ;
- Delegação (*delegation*): ação de confiar.

De acordo com os autores, o termo confiança não pode ser confundido com delegação. A delegação é necessariamente uma ação, o resultado de uma decisão. Além disto, ela cria uma relação entre as entidades  $x$  e  $y$  de acordo com a ação  $a$ . A confiança é "o *background* mental para a delegação"(CASTELFRANCHI; FALCONE, 2001).

De acordo com o modelo, as crenças para a confiança são as apresentadas na tabela 1. Quando um agente intencional estiver em uma situação com delegação fraca, aquelas em que não pressupõe nenhum acordo ou promessa, as crenças apresentadas na tabela 2 são necessárias. Quando a situação for em delegação forte, aquelas em que o agente confiado  $y$  tem ciência que o agente  $x$  precisa/depende dele, é criada a crença motivacional: o agente  $x$  acredita que  $y$  tem motivos para ajudá-lo. O lado afetivo da confiança não é explorado no trabalho dos autores.

Tabela 1 – Tipos de crenças utilizadas para confiança.

Tipo	Descrição
Competência	O agente deve acreditar que $y$ pode produzir o resultado.
Disposição	$y$ irá fazer o solicitado.
Dependência	O agente acredita que ele precisa, depende, ou pelo menos é melhor depender de $y$ .
Realização	O objetivo $g$ será realizado com a ajuda de $y$ .

Tabela 2 – Tipos de crenças para delegação fraca.

Tipo	Descrição
Complacência	O agente acredita que $y$ decidiu e irá executar a ação $a$ .
Persistência	O agente acredita que $y$ está comprometido com a ação $a$ .
Auto-confiança	O agente acredita que $y$ acredita que pode executar a ação $a$ .

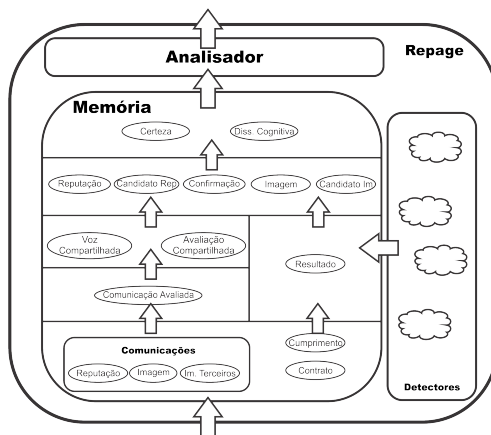
### 2.4.1.2 Repage

Repage (SABATER; PAOLUCCI; CONTE, 2006) <sup>1</sup> é um modelo baseado em uma teoria cognitiva de reputação, projetado para ser adicionado à arquiteturas de agentes. Dois conceitos importantes no modelo são: *imagem* e *reputação*. Imagem é uma crença avaliativa que diz se o alvo é bom ou ruim de acordo com uma habilidade, padrão ou norma. Reputação é uma meta-crença sobre a existência de uma avaliação, que de certa forma é impessoal, comunicada.

Imagem e reputação são avaliações sociais. Uma avaliação social é feita por um agente, avaliando outra entidade, de acordo com uma habilidade, padrão, ou norma. Ela é composta por três elementos: o agente avaliado; o papel representando o objeto da avaliação; e uma distribuição de probabilidade feita sobre um conjunto de rótulos (PINYOL; SABATER-MIR, 2009).

A arquitetura do modelo Repage é composta por três elementos principais: memória (*memory*), um conjunto de detectores (*detectors*) e pelo analisador (*analyzer*). Estes componentes estão representados na figura 2.

Figura 2 – Arquitetura do modelo Repage.



Fonte: Adaptado de Sabater, Paolucci e Conte (2006).

A memória do Repage é formada por um conjunto de referências aos predicados da memória de uso geral do agente. Desta forma, uma mudança em um destes predicados é visível para ambas as memórias. No Repage,

<sup>1</sup>REputation and ImAGE



predicados são objetos que contêm avaliações sociais. Além disto, predicados têm um conjunto de antecedentes e consequentes.

A memória é dividida em cinco níveis:

1. No primeiro nível estão os contratos (*contracts*), os resultados dos contratos (*fulfilments*) e as comunicações de informantes (*informers' communication*);
2. No segundo são definidos dois predicados, um para inferir novos resultados a partir de um contrato e seus resultados, e o outro para ponderar e valorar avaliações obtidas por comunicação;
3. No terceiro nível estão os predicados de voz compartilhada (*shared voice*) e avaliação compartilhada (*shared evaluation*). O primeiro é o elemento principal para construir reputação, e o segundo para construir imagem;
4. O nível seguinte é formado por predicados para imagem candidata (*candidate image*), reputação candidata (*candidate reputation*), imagem (*image*), reputação (*reputation*) e confirmação (*confirmation*). Os dois primeiros são candidatos por não possuírem suporte suficiente para serem imagem e reputação. O predicado confirmação representa a qualidade de informações prévias;
5. O último nível apresenta dois predicados: dissonância cognitiva (*cognitive dissonance*) e certeza (*certainty*). O primeiro refere-se a informações contraditórias que são relevantes. O segundo representa informações que são totalmente confiáveis.

Detectores são unidades de inferência especializadas em um grupo de predicados. Eles são responsáveis por inferir novos predicados a partir de predicados existentes, por remover predicados desnecessários e por formar a rede de dependência entre os predicados. O analisador é responsável em propor ações que podem melhorar a precisão dos predicados na memória do RePage e para resolver dissonâncias cognitivas visando produzir uma situação de certeza.

#### 2.4.1.3 ReGreT

ReGreT (SABATER; SIERRA, 2001; SABATER-MIR, 2002) é um modelo numérico de reputação, onde reputação é definida como uma “opinião ou visão de alguém sobre algo”. Inicialmente, o modelo foi proposto para ambientes de comércio eletrônico onde o papel das relações sociais é

importante, porém sua estrutura permite que seja utilizado em outros ambientes. O cálculo da reputação utiliza informações de experiências diretas, informações de terceiros e estrutura social.

A reputação é dividida em três dimensões: *individual*, *social* e *ontológica*. Com o objetivo de descrever melhor estes três componentes, grupos de agentes são denotados por letras maiúsculas ( $\mathcal{A}, \mathcal{B}, \dots$ ) e agentes pertencentes a estes grupos por letras minúsculas ( $a_1, a_2 \in \mathcal{A}, b_1, b_2 \in \mathcal{B}, \dots$ ). O conjunto de identificadores de agentes é denotado por  $A$ , e o conjunto de resultados de diálogos entre dois agentes por  $O$ , onde um resultado (*outcome*)  $o$  é composto pelo contrato inicial e o pelo resultado da transação.

De acordo com a dimensão individual, o agente avaliador  $a$  só considera suas próprias experiências para o cálculo da reputação (confiança direta) sobre um agente avaliado  $b$ . A dimensão social analisa as experiências do integrante  $a_i$  pertencente ao grupo  $\mathcal{A}$ , em que o agente  $a$  faz parte, avaliando o agente  $b$  e os integrantes pertencentes ao grupo  $\mathcal{B}$ , em que  $b$  está inserido. A dimensão ontológica permite que valores de confiança e reputação de múltiplos aspectos sejam combinados visando o cálculo de comportamentos mais complexos. Por exemplo, a reputação em aspectos como *disponibilidade do conteúdo* e *qualidade do material* podem ser combinadas para definir uma *boa plataforma de cursos online*.

Os autores utilizam o conceito de *impressão*. Uma impressão  $\iota$  é uma “avaliação subjetiva feita por um agente sobre um certo aspecto de um resultado” (SABATER; SIERRA, 2001). Ela é da forma

$$\iota = (a, b, o, \varphi, t, W),$$

onde  $a, b \in A$  são os agentes que dialogaram, sendo que  $a$  está julgando  $b$ ,  $o \in O$  é o resultado da interação,  $\varphi$  é a variável de  $o$  sendo avaliada,  $t$  é o tempo em que a impressão foi registrada e  $W \in [-1, 1]$  é a avaliação dada a variável do resultado avaliado. Por exemplo, o resultado

$$o_a = ($$

$$\text{DuraçãoDoCurso}_c = \text{trêsSemanas} \wedge \text{Preço}_c = 500 \wedge$$

$$\text{QualidadeDoMaterial}_c = \text{muitoBoa} \wedge$$

$$\text{DuraçãoDoCurso} = \text{umaSemana} \wedge \text{Preço} = 500 \wedge$$

$$\text{QualidadeDoMaterial} = \text{ruim})$$

de acordo com a perspectiva do agente  $a$  pode ser interpretado como: A duração do curso combinada ( $\text{DuraçãoDoCurso}_c$ ) era de *trêsSemanas*, porém o curso durou *umaSemana*; o preço combinado ( $\text{Preço}_c$ ) correspondeu com preço pago ( $\text{Preço}$ ). A qualidade esperada pelo agente  $a$  era *MuitoBoa*, po-

rém a qualidade, de acordo com o agente é *ruim*. Com isto, uma *impressão* para o resultado  $o_a$  pode ser

$$(a, b, o_a, DuraçãoDoCurso, 10/12/2015, -0.70),$$

ou seja, o agente  $a$ , registrou uma impressão sobre o agente  $b$ , considerando o atributo *DuraçãoDoCurso* do resultado  $o_a$ , no tempo 10/12/2015, avaliando subjetivamente como  $-0,70$ .

O agente possui uma base de dados de impressões (*IDB*), a qual é utilizada para o cálculo da reputação individual. A reputação individual que um agente  $a$  tem sobre um agente  $b$ , segundo algum aspecto  $\varphi$ , respeitando o padrão  $p$  até o tempo  $t$ , é definida como

$$R^t(IDB_p^a) = \sum_{t_i \in IDB_p^a} \rho(t, t_i) \cdot W_i, \quad (2.2)$$

onde

$$\rho(t, t_i) = \frac{f(t_i, t)}{\sum_{t_j \in IDB_p^a} f(t_j, t)} \quad (2.3)$$

é uma função que atribui pesos maiores para impressões mais recentes.  $f(t_i, t)$  é uma função dependente do tempo. Como exemplo, o autor define  $f(t_i, t) = \frac{t_i}{t}$ . Como  $W_i \in [-1, 1]$  e  $\rho(t, t_i)$  é um valor normalizado,  $R^t(IDB_p^a) \in [-1, 1]$ . Daqui para frente, a notação  $R_{a \rightarrow b}(\varphi)$  é utilizada para representar  $R^t(IDB_p^a)$ . Para possibilitar isso, o padrão  $p$  aplicado sobre *IDB*, é assumido como  $p = (a, b, \_, \varphi, \_, \_)$ : *true*, ou seja, não é definido nenhum padrão que seleciona um subconjunto das impressões do agente  $a$  sobre o agente  $b$ . Além disso, o tempo  $t$  é assumido como o atual.

A qualidade do valor da reputação é definida a partir da confiabilidade. Na reputação individual, a confiabilidade é computada considerando o número de impressões utilizadas e a variabilidade dos resultados  $W_i$ . Conforme o número de impressões aumenta, a suspeita do resultado diminui. Existe um número máximo de impressões, de modo que, a partir dele os agentes têm *intimidade* (*itm*), onde o grau de confiabilidade atinge o valor máximo. A função 2.4, onde  $|IDB_p^a|$  é a cardinalidade de  $IDB_p^a$ , pode ser utilizada para modelar a influência que o número de impressões utilizadas,  $Ni(IDB_p^a)$ , tem na definição da confiabilidade.

$$Ni(IDB_p^a) = \begin{cases} \sin\left(\frac{\pi}{2 \cdot itm} |IDB_p^a|\right) & \text{se } |IDB_p^a| \in [0, itm] \\ 1 & \text{caso contrário} \end{cases} \quad (2.4)$$

A variabilidade nos resultados,  $Dt(IDB_p^a)$ , define a credibilidade do

valor da reputação individual. Ela é definida pela função 2.5, ficando no intervalo  $[0, 1]$ . Valores próximos a 1 indicam alta variabilidade dos valores, ou seja, baixa credibilidade, e valores próximos a 0 indicam baixa variabilidade e alta credibilidade.

$$Dt(IDB_p^a) = \sum_{t_i \in IDB_p^a} \rho(t, t_i) \cdot |W_i - R^t(IDB_p^a)| \quad (2.5)$$

Assim, a confiabilidade da reputação individual é definida por

$$RL(IDB_p^a) = Ni(IDB_p^a) \cdot (1 - Dt(IDB_p^a)). \quad (2.6)$$

A notação  $RL_{a \rightarrow b}(\varphi)$  é utilizada a partir de agora para representar  $RL^t(IDB_p^a)$ .

A segunda dimensão da reputação, dimensão social, permite a modelagem de relações entre grupos. Ela divide-se em três fontes de informação:

1. O agente avaliador  $a$  interagindo com o grupo  $\mathcal{B}$  do agente avaliado

$$R_{a \rightarrow \mathcal{B}}(\varphi) = \sum_{b_i \in \mathcal{B}} \omega^{ab_i} \cdot R_{a \rightarrow b_i}(\varphi) \quad (2.7)$$

onde  $\sum_{b_i \in \mathcal{B}} \omega^{ab_i} = 1$  e  $\omega^{ab_i}$  permite que o agente  $a$  dê peso maior para agentes próximos ao agente  $b$ . Na incerteza da existência de agentes próximos, a função  $\omega^{ab_i} = \frac{1}{|\mathcal{B}|}$  é utilizada para atribuir peso igual a todos agentes de  $\mathcal{B}$ . A confiabilidade desta reputação é definida como

$$RL_{a \rightarrow \mathcal{B}}(\varphi) = \sum_{b_i \in \mathcal{B}} \omega^{ab_i} \cdot RL_{a \rightarrow b_i}(\varphi). \quad (2.8)$$

2. O grupo  $\mathcal{A}$  do agente avaliador  $a$  avaliando o agente  $b$

$$R_{\mathcal{A} \rightarrow b}(\varphi) = \sum_{a_i \in \mathcal{A}} \omega^{a_i b} \cdot R_{a_i \rightarrow b}(\varphi), \quad (2.9)$$

$$RL_{\mathcal{A} \rightarrow b}(\varphi) = \sum_{a_i \in \mathcal{A}} \omega^{a_i b} \cdot RL_{a_i \rightarrow b}(\varphi) \quad (2.10)$$

onde  $\sum_{a_i \in \mathcal{A}} \omega^{a_i b} = 1$ .

3. O grupo  $\mathcal{A}$  do agente  $a$  avaliando o grupo  $\mathcal{B}$  do agente  $b$

$$R_{\mathcal{A} \rightarrow \mathcal{B}}(\varphi) = \sum_{a_i \in \mathcal{A}} \omega^{a_i \mathcal{B}} \cdot R_{a_i \rightarrow \mathcal{B}}(\varphi), \quad (2.11)$$

$$RL_{\mathcal{A} \rightarrow \mathcal{B}}(\varphi) = \sum_{a_i \in \mathcal{A}} \omega^{a_i \mathcal{B}} \cdot RL_{a_i \rightarrow \mathcal{B}}(\varphi). \quad (2.12)$$

$$\text{onde } \sum_{a_i \in \mathcal{A}} \omega^{a_i \mathcal{B}} = 1.$$

$\omega^{a_i b}$  e  $\omega^{a_i \mathcal{B}}$ , dos itens 2 e 3, são utilizados para dar mais importância para as informações provenientes dos agentes com mais credibilidade ou com maior posição hierárquica.

A reputação final  $SR_{a \rightarrow b}(\varphi)$  feita pelo agente  $a$  para com o agente  $b$ , considerando a dimensão individual e social é definida como

$$SR_{a \rightarrow b}(\varphi) = \xi_{ab} \cdot R_{a \rightarrow b}(\varphi) + \xi_{a\mathcal{B}} \cdot R_{a \rightarrow \mathcal{B}}(\varphi) + \xi_{\mathcal{A}b} \cdot R_{\mathcal{A} \rightarrow b}(\varphi) + \xi_{\mathcal{A}\mathcal{B}} \cdot R_{\mathcal{A} \rightarrow \mathcal{B}}(\varphi) \quad (2.13)$$

e a confiabilidade deste valor é definida como

$$SRL_{a \rightarrow b}(\varphi) = \xi_{ab} \cdot RL_{a \rightarrow b}(\varphi) + \xi_{a\mathcal{B}} \cdot RL_{a \rightarrow \mathcal{B}}(\varphi) + \xi_{\mathcal{A}b} \cdot RL_{\mathcal{A} \rightarrow b}(\varphi) + \xi_{\mathcal{A}\mathcal{B}} \cdot RL_{\mathcal{A} \rightarrow \mathcal{B}}(\varphi) \quad (2.14)$$

onde  $\xi_{ab} + \xi_{a\mathcal{B}} + \xi_{\mathcal{A}b} + \xi_{\mathcal{A}\mathcal{B}} = 1$ . Estes parâmetros permitem que o agente ajuste, dinamicamente ou não, a influência que cada fonte de informação irá exercer.

A terceira dimensão do modelo, dimensão ontológica, é representada com grafos, onde cada nodo representa a reputação de acordo com um aspecto e a combinação de nodos gera uma reputação global. Assim, a confiança e sua confiabilidade, são definidas, de acordo com a dimensão ontológica, respectivamente, pelas funções 2.15 e 2.16.

$$OR_{a \rightarrow b}(i) = \sum_{j \in \text{filho}(i)} \omega_{ij} \cdot OR_{a \rightarrow b}(j) \quad (2.15)$$

$$ORL_{a \rightarrow b}(i) = \sum_{j \in \text{filho}(i)} \omega_{ij} \cdot ORL_{a \rightarrow b}(j) \quad (2.16)$$

onde  $OR_{a \rightarrow b}(j) = SR_{a \rightarrow b}(j)$  quando  $j$  é um aspecto atômico.

Por exemplo, assumindo que a reputação como *plataforma de curso online* é composta pelas reputações de *disponibilidade do conteúdo* e *qualidade do material*. Supondo que o agente  $a$  avaliou a reputação social do agente  $b$  considerando a disponibilidade do conteúdo com 0.5, e a qualidade do material com 0.8, a reputação como plataforma de curso online pode ser

definida, pelo agente  $a$ , considerando a dimensão ontológica como

$$OR_{a \rightarrow b}(\text{plataformaOnline}) = 0,7 \cdot 0,5 + 0,3 \cdot 0,8.$$

onde 0,7 e 0,3 representam, respectivamente, os pesos da reputação quanto a disponibilidade do conteúdo e quanto a qualidade do material na reputação como plataforma de curso online.

#### 2.4.1.4 FIRE

FIRE (HUYNH; JENNINGS; SHADBOLT, 2006)<sup>2</sup> é um modelo numérico de confiança e reputação, projetado para SMAs abertos. O modelo incorpora confianças de interação e baseada em papel, e reputações de testemunhas e certificada.

A confiança de interação é computada a partir do conjunto de interações entre dois agentes. Este componente é implementado da mesma forma como a reputação individual do modelo ReGreT descrito na seção 2.4.1.3.

A confiança baseada em papel é determinada a partir de papéis dos dois agentes. Regras são utilizadas para determinar o valor da confiança. Elas são da forma

$$\text{regra} = (\text{papel}_x, \text{papel}_y, c, v_D, e_D)$$

onde  $\text{papel}_x$  e  $\text{papel}_y$  são papéis dos agentes  $x$  e  $y$ ,  $v_D \in [-1, 1]$  é o desempenho esperado do agente  $y$  em uma interação com  $x$ , em relação ao termo  $c$ ,  $e_D \in [0, 1]$  é o nível padrão de influência desta regra no valor da confiança baseada em papel. O valor final da confiança baseada em papel é definido como

$$\tau_P(x, y, c) = \frac{\sum_{\text{regra}_i \in \text{Regras}(x, y, c)} e_{D_i} \cdot v_{D_i}}{\sum_{\text{regra}_i \in \text{Regras}(x, y, c)} e_{D_i}},$$

onde  $\text{regra}_i = (\text{papel}_x, \text{papel}_y, c, v_{D_i}, e_{D_i})$  é uma regra do conjunto de regras  $\text{Regras}(x, y, c)$ , que por sua vez é um subconjunto com as regras importantes para os papéis de  $x$  e  $y$  e para o termo  $c$  selecionado. A confiabilidade deste valor é especificada pelo dono do agente.

A reputação de testemunhas é determinada a partir de informações fornecidas por outros agentes do sistema. Ela é computada como

---

<sup>2</sup>FIRE: 'Fides' (confiança em latim) e REputation

$$\tau_T(x, y, c) = \sum_{r_i \in R_T(x, y, c)} \omega(r_i) \cdot v_i,$$

onde  $R_T(x, y, c)$  é o conjunto de avaliações fornecidas por testemunhas ao agente  $x$ , o peso  $\omega(r_i)$  é o peso dado a avaliação  $r_i$ , dando mais importância para avaliações mais recentes e  $v_i$  é o valor da avaliação de  $r_i$ . A confiabilidade deste valor é calculada a partir do número de interações e do desvio destes valores. Esta confiança também é similar a proposta do modelo Re-GreT.

A reputação certificada é computada a partir de avaliações que a entidade avaliada  $y$  fornece sobre si mesma. Estas avaliações são obtidas por  $y$  por meio de operações que ela faz com outras entidades do sistema. O cálculo da reputação certificada e da confiabilidade é igual ao feito pela reputação de testemunha. Sendo que a entrada é o conjunto de avaliações fornecidas por  $y$ , ao invés de informações fornecidas por testemunhas.

O valor final da confiança é obtido a partir da combinação das confianças e reputações. Assim, a confiança final é definida como

$$\tau(x, y, c) = \frac{\sum_{k \in \{I, P, T, C\}} w_k \cdot \tau_k(x, y, c)}{\sum_{k \in \{I, P, T, C\}} w_k},$$

e a confiabilidade como

$$\rho \tau(x, y, c) = \frac{\sum_{k \in \{I, P, T, C\}} w_k}{\sum_{k \in \{I, P, T, C\}} W_k}$$

onde

$$w_k = W_k \cdot \rho \tau_k(x, y, c)$$

e  $W_I$ ,  $W_P$ ,  $W_T$  e  $W_C$  são os coeficientes correspondentes aos componentes de Interação, Papel, Testemunho e Certificada. O usuário define quais são os valores destes coeficientes.

## 2.5 EMOÇÕES

A computação afetiva, termo definido por Picard (PICARD; PICARD, 1997), é a “computação que se relaciona com, surge de, ou deliberadamente influencia emoções”.<sup>3</sup> A introdução de emoções em sistemas computacionais objetiva melhorar a interação humano-máquina; testar, refinar e desenvolver hipóteses emocionais; colaborar no refinamento de técnicas de Inteligência

---

<sup>3</sup>Nossa tradução

Artificial (IA) (MARSELLA; GRATCH; PETTA, 2010).

Modelos computacionais de emoções são

Sistemas de software designados a fornecer à agentes autônomos mecanismos próprios para processar informações de emoções, elicitacão de emoções sintéticas, e geraçao de comportamentos emocionais (RODRÍ-GUEZ; RAMOS, 2014).

A criaçao de modelos computacionais de emoções é fundamentada em teorias de emoções. Nas teorias apreciativas (*appraisal theories*), a elicitacão e diferenciaçao de emoções são feitas por meio da avaliaçao, ou apreciaçao subjetiva do indivíduo considerando uma situaçao, objeto ou evento (SCHE-RER, 1999), ou seja, analisando entidades discretas. Em teorias dimensionais, emoções e outros fenômenos afetivos são visualizados como pontos em um espaço dimensional contínuo (MARSELLA; GRATCH; PETTA, 2010). Embora existam outras, estas duas são as mais comuns em modelos computacionais de emoções, tais como WASABI (BECKER-ASANO, 2008), FLAME (EL-NASR; YEN; IOERGER, 2000), ALMA (GEBHARD, 2005) e EMA (MARSELLA; GRATCH, 2009).

O modelo OCC, proposto por Ortony, Clore e Collins (1990), é um modelo de teorias apreciativas frequentemente utilizado. Ele classifica 22 emoções em três grupos: baseadas em eventos, baseadas em agentes e baseadas em aspectos de objetos (ORTONY; CLORE; COLLINS, 1990).

Um dos representantes mais significativos de teorias dimensionais é o modelo de estados emocionais PAD (*Pleasure-Arousal-Dominance*). Ele é composto por três eixos independentes e bipolares representando prazer (*pleasure*), ativaçao (*arousal*) e dominância (*dominance*). O primeiro componente, prazeroso / não-prazeroso, corresponde aos julgamentos avaliativos, quanto maior a avaliaçao, mais prazeroso. O segundo componente, ativaçao / desativaçao, representa o nível de atividade. O terceiro componente, dominância / submissao, é uma medida de controle, ou poder (RUSSELL; MEHRABIAN, 1977; MEHRABIAN, 1996).

Outro campo de estudo sobre emoções e IA, está em como integrar emoções em mecanismos de tomada de decisao. Em especial, em arquiteturas de agentes cognitivos, como BDI, onde a influencia de emoções emerge em vários componentes do agente, como por exemplo, nas observaçoes, em crenças, na escolha das intenções (JIANG; VIDAL; HUHNS, 2007; CÔRTEZ, 2013; PUICĂ; FLOREA, 2013; LEJMI-RIahi; KEBAIR; SAID, 2014).

Segundo Jiang, Vidal e Huhns (2007), para incorporar emoções em agentes é necessário responder a três questões. A seguir são apresentadas estas questões e suas respostas no contexto desta pesquisa.

### 1. Como medir ou apresentar emoções?



Emoções são medidas segundo o modelo WASABI descrito na seção 2.5.1;

2. Como emoções afetam o processo de tomada de decisão?  
No modelo proposto, emoções influenciam na escolha de entidades para interagir. A influência direta está em crenças que são utilizadas para definir a confiança em cada entidade;
3. Como atualizar o estado de emoções?  
O estado emocional é atualizado a partir de impulsos emocionais positivos ou negativos. Estes impulsos são provocados por percepções, intenções ou crenças do agente.

Dentre os modelos computacionais de emoções encontrados na literatura, dois são apresentados, um com maior influência de teorias apreciativas, abordagem de Bitencourt, Silveira e Marchi (2013), e outro com maior influência de teorias dimensionais, WASABI (BECKER-ASANO, 2008; BECKER-ASANO; WACHSMUTH, 2010; BECKER-ASANO, 2014). Estes modelos são apresentados visando realizar o objetivo de analisar modelos computacionais de emoções.

### 2.5.1 WASABI

O modelo de emoções WASABI<sup>4</sup> (BECKER-ASANO, 2008; BECKER-ASANO; WACHSMUTH, 2010; BECKER-ASANO, 2014), considera emoções como “estados atuais com qualidade e intensidade específicas”, resultantes de um processo de comunicação, com valência positiva ou negativa, e durando significativamente menos que humor. Humor é modelado como “um estado em segundo plano com uma qualidade afetiva muito mais simples que emoções”. Embora possuam significado diferente, emoções e humor influenciam um no outro.

O modelo simula nove emoções primárias: raiva (*angry*), irritado (*annoyed*), entediado (*bored*), deprimido (*depressed*), medroso (*fearful*), feliz (*happy*), triste (*sad*), surpreso (*surprised*) e concentrado (*concentrated*), este último não sendo um estado emocional; e três emoções secundárias: esperança (*hope*), medo confirmado (*fears-confirmed*) e alívio (*relief*).

Segundo este modelo, o agente tem um estado homeostático. Sempre que um estímulo emocional, interno ou externo, é detectado, a sua valência serve como um impulso que perturba este estado homeostático do “equilíbrio emocional” do agente, causando certos níveis de prazer e ativação no módulo

---

<sup>4</sup> Acrônimo de WASABI *Affect Simulation for Agents with Believable Interactivity*

emocional. Além disto, um processo dinâmico começa, objetivando a volta do equilíbrio emocional do agente.

A consciência do agente de suas emoções primárias e secundárias é definida de forma separada. Considerando a probabilidade da consciência de emoções primárias, seja  $pe$  uma emoção primária,  $d_{pe}$  a distância entre o valor atual PAD e o valor PAD da emoção,  $\phi_{pe}$  o limiar de ativação e  $\Delta_{pe}$  o limiar de saturação. O cálculo da probabilidade da consciência de emoções  $w_{pe}$  inicia quando  $d_{pe}$  fica menor que  $\phi_{pe}$ , e é executado até que  $d_{pe}$  fique abaixo de  $\Delta_{pe}$ , ou seja, ela continua até que  $w_{pe}$  fique igual a intensidade base  $i_{pe}$  da emoção. Uma emoção que possui a  $i_{pe} = 0$  precisa ser desencadeada por cognição antes de ganhar um valor  $w_{pe}$  diferente de zero. Para emoções com mais do que uma representação no espaço PAD, é considerada a menor distância ao ponto referenciado. A equação 2.17 apresenta a determinação de  $w_{pe}$ .

$$w_{pe} = \left(1 - \frac{d_{pe} - \Delta_{pe}}{\phi_{pe} - \Delta_{pe}}\right) \cdot i_{pe} \text{ com } \phi_{pe} > \Delta_{pe} \quad \forall pe \in \{pe_1, \dots, pe_9\} \quad (2.17)$$

A probabilidade da consciência de qualquer emoção secundária assume que a sua intensidade base é, por padrão, configurada como 0. Toda emoção secundária precisa ser desencadeada por um processo cognitivo. Outra propriedade de emoções secundárias é que ela tem um parâmetro para seu “tempo de vida”, com uma função de decaimento.

## 2.5.2 Steunebrink e TrustE

Com base nos trabalhos de Steunebrink, Dastani e Meyer (2007, 2009), Bitencourt, Silveira e Marchi (2013) criaram um modelo em que oito emoções do modelo OCC são definidas. Quatro disparadas por eventos: esperança, medo, alegria e angústia; e quatro disparadas por ações: orgulho, vergonha, admiração e reprovação. Elas podem ser positivas (alegria, esperança, orgulho e admiração), ou negativas (angústia, medo, vergonha e reprovação).

Sejam  $a$  e  $b$  agentes,  $E$  um evento,  $A$  uma ação, *positivo* uma resposta positiva em relação a alguma coisa, e *negativo* uma resposta negativa em relação a alguma coisa, são utilizadas as seguintes proposições (BITENCOURT; SILVEIRA; MARCHI, 2013):

- $Ação_a(A)$ : uma ação  $A$  realizada por  $a$ ;
- $Provável(E)$ : um evento  $E$  que possivelmente irá ocorrer;

- $Atual(E)$ : um evento  $E$  que está ocorrendo;
- $Satisfeito_a(E)$ :  $a$  é positivo sobre a consequência de  $E$ ;
- $Insatisfeito_a(E)$ :  $a$  é negativo sobre a consequência de  $E$ ;
- $Aprova_a(A)$ :  $a$  é positivo sobre a ação  $A$  de algum agente;
- $Desaprova_a(A)$ :  $a$  é negativo sobre a ação  $A$  de algum agente;

Estas proposições são utilizadas para definir emoções. Primeiro são apresentadas as quatro emoções disparadas por eventos, e em seguida as quatro disparadas por ações.

$$Esperança_a(E) \leftrightarrow Satisfeito_a(E) \wedge Provável(E) \quad (2.18)$$

Um agente sente esperança se ele está satisfeito em relação a provável ocorrência de um evento.

$$Medo_a(E) \leftrightarrow Insatisfeito_a(E) \wedge Provável(E) \quad (2.19)$$

Um agente sente medo se ele está insatisfeito em relação a provável ocorrência de um evento.

$$Alegria_a(E) \leftrightarrow Satisfeito_a(E) \wedge Atual(E) \quad (2.20)$$

Um agente sente alegria quando está satisfeito em relação a um evento atual.

$$Angústia_a(E) \leftrightarrow Insatisfeito_a(E) \wedge Atual(E) \quad (2.21)$$

Um agente sente angústia quando está insatisfeito em relação a um evento atual.

$$Orgulho_a(A, a) \leftrightarrow Aprova_a(A) \wedge Ação_a(A) \quad (2.22)$$

Um agente sente orgulho se ele aprova sua própria ação.

$$Vergonha_a(A, a) \leftrightarrow Desaprova_a(A) \wedge Ação_a(A) \quad (2.23)$$

Um agente sente vergonha se ele desaprova sua própria ação.

$$Admiração_a(A, b) \leftrightarrow Aprova_a(A) \wedge Ação_b(A) \quad (2.24)$$

Um agente sente admiração por outro agente, se ele aprova a ação desse agente.

$$\text{Reprova\c{c}o}_a(A,b) \leftrightarrow \text{Desaprova}_a(A) \wedge \text{A\c{c}o}_b(A) \quad (2.25)$$

Um agente sente reprova\c{c}o por outro agente, se ele desaprova a a\c{c}o deste agente.

A intensidade das emo\c{c}o\c{e}s \c{e} calculada conforme a equa\c{c}o 2.26.

$$I(E,t) = \max(0, P(E,t) - L(E,t)), \quad (2.26)$$

onde  $I(E,t)$  \c{e} a fun\c{c}o de intensidade da emo\c{c}o  $E$  no tempo  $t$ ,  $P(E,t)$  \c{e} a fun\c{c}o de *potencialidade* de  $E$  em  $t$ , e  $L(E,t)$  a fun\c{c}o do *limiar* de  $E$  em  $t$ .

A potencialidade de uma emo\c{c}o representa a possibilidade dela ocorrer, ou seja, quanto maior a potencialidade de uma emo\c{c}o, mais prov\u00e1vel que ela ocorra. Ela \c{e} apresentada na equa\c{c}o 2.27. O limiar representa o valor m\u00ednimo para que a emo\c{c}o tenha alguma intensidade. Ele \c{e} apresentado na equa\c{c}o 2.28.

$$P(E,t) = \frac{\sum_{i=1}^N VI_i}{N}, \quad (2.27)$$

onde  $VI_i \in [0, 1]$  (as “vari\u00e1veis de intensidade”) representam os valores das vari\u00e1veis que afetam a intensidade da emo\c{c}o  $E$ .

$$L(E,t) = \begin{cases} 1 - H(t), & \text{se } E \text{ \c{e} uma emo\c{c}o positiva;} \\ H(t), & \text{caso } E \text{ seja uma emo\c{c}o negativa.} \end{cases} \quad (2.28)$$

$H(t) \in [0, 1]$  \c{e} uma fun\c{c}o que representa o humor do agente no tempo  $t$ .

Neste cap\u00edtulo os temas *agente, ambiente, sistema multi-contexto, confian\c{c}a e emo\c{c}o\c{e}s* foram introduzidos. Foram descritos quatro modelos de confian\c{c}a, dois cognitivos e dois num\u00e9ricos, visando realizar o objetivo de analisar modelos de confian\c{c}a. Dois modelos de emo\c{c}o\c{e}s foram apresentados: (i) a abordagem de Bitencourt, Silveira e Marchi (2013), a qual relaciona o modelo com confian\c{c}a; e (ii) o modelo WASABI (BECKER-ASANO, 2008; BECKER-ASANO; WACHSMUTH, 2010; BECKER-ASANO, 2014), utilizado no modelo proposto. E com isto o objetivo de analisar modelos computacionais de emo\c{c}o\c{e}s foi concretizado.

### 3 TRABALHOS RELACIONADOS

Neste capítulo são apresentados os principais trabalhos que, de alguma forma, se assemelham com a proposta desta pesquisa. São apresentadas abordagens para integrar confiança em agentes BDI, definições de agentes como sistemas multi-contexto, e trabalhos que relacionam emoções e confiança.

#### 3.1 GRADED BDI

Casali, Godo e Sierra (2005) apresentam um modelo de agente como SMC com crenças, desejos e intenções graduadas. A graduação nas crenças representa quanto o agente acredita que uma fórmula é verdadeira. Desejos podem ser positivos ou negativos. A graduação em desejos positivos representa o nível de preferência do agente sobre o desejo, enquanto que a graduação negativa representa o nível de rejeição sobre ele. Graus em intenções modelam a relação custo/benefício em atingir um objetivo.

O contexto de crenças (BC) é construído a partir de lógica proposicional clássica estendida com modalidade de ações. Assim, a interpretação de  $[\alpha]A$  é que após a execução de  $\alpha$ ,  $A$  será verdade. Para raciocinar sobre a credibilidade das crenças, é utilizada lógica infinitamente valorada de Łukasiewicz (ŁUKASIEWICZ, 1920; HAY, 1963). Enquanto que uma proposição Crisp possui dois valores verdade, na lógica infinitamente valorada de Łukasiewicz, existem infinitos possíveis valores verdade. Assim, fórmulas de BC podem ser *Crisp* (*Não modais*), ou *B-modais*, onde  $B$  é um operador modal unário fuzzy. O contexto de crenças interage com os contextos de desejos, planejamento e comunicação.

O contexto de desejos (DC) é composto por desejos positivos e negativos. Os positivos correspondem aos desejos que o agente gostaria que fossem reais. Os negativos representam as suas rejeições, as coisas que ele não gostaria que se tornassem reais. Ambos podem ser graduados. Desta forma, a linguagem de DC é definida como uma extensão de uma linguagem proposicional clássica.

$D^+ \varphi$  é lido como “ $\varphi$  é positivamente desejado”, e o seu grau de verdade representa o nível de satisfação do agente em  $\varphi$  se tornar verdade.  $D^- \varphi$  é lido como “ $\varphi$  é negativamente desejado”, e o seu grau de verdade representa o nível de desgosto do agente em  $\varphi$  se tornar verdade. Assim como em crenças, é utilizada lógica infinitamente valorada de Łukasiewicz. Esta é estendida com o conectivo  $\Delta$ . Para qualquer modal  $\phi$ , se  $\phi$  tem valor  $< 1$  então  $\Delta\phi$  recebe valor 0; caso contrário, se  $\phi$  tem valor 1 então  $\Delta\phi$  também recebe

valor 1. Fórmulas DC podem ser crisp ou infinitamente valoradas (modais). O contexto de desejos interage com intenções, crenças e planeamento.

Uma intenção representada no contexto de intenções (IC) é considerada uma *pro-attitude* fundamental com representação explícita. *Pro-attitudes* guiam as ações do agente (WOOLDRIDGE, 2009). Intenções não podem depender só dos benefícios, ou satisfação, da realização de um objetivo  $\varphi$ , mas também do estado  $w$  do mundo, e dos custos envolvidos na transformação em um mundo  $w_i$  onde a fórmula  $\varphi$  é verdadeira. Ao adicionar graus em intenções, é possível medir os custos/benefícios envolvidos nas ações do agente rumo ao objetivo. Intenções dependem do conhecimento do agente sobre o mundo. Este contexto parte de uma linguagem proposicional básica, como no contexto de crenças, sendo adicionado o operador modal  $I$ . A esta linguagem, é adicionada a lógica infinitamente valorada de Łukasiewicz para representar os graus das intenções. A intenção de tornar  $\varphi$  verdade deve ser a consequência de encontrar um plano  $\alpha$  factível, o qual permite alcançar um estado do mundo onde  $\varphi$  é válido. O contexto de intenções interage com os contextos de planeamento, comunicação e desejos.

O contexto de comunicação (CC) é construído a partir de linguagem de primeira ordem restrita à cláusulas de Horn. É caracterizada pela interface única e bem definida com o ambiente. Interage com o ambiente, intenções, planeamento e crenças.

O contexto de planeamento (PC) contrói planos que permitem ao agente se mover do seu estado atual do mundo para outro, onde uma dada fórmula é satisfeita. A linguagem deste contexto é de primeira ordem restrita a cláusulas de Horn. A teoria de planeamento inclui três predicados especiais:

1.  $action(\alpha, P, A, c_\alpha)$  onde  $\alpha \in \Pi_0$  é uma ação elementar pertencente ao conjunto  $\Pi_0$  de ações elementares,  $P$  é o conjunto de precondições;  $A$  são as pós-condições e  $c_\alpha \in [0, 1]$  é o custo normalizado da ação.
2.  $plan(\varphi, \alpha, P, A, c_\alpha, r)$  onde  $\alpha \in \Pi$  é uma ação composta representando o plano para realizar o desejo  $\varphi$ ,  $P$  são as precondições de  $\alpha$ ,  $A$  são as pós-condições  $\varphi \in A$ ,  $c_\alpha$  é o custo normalizado de  $\alpha$  e  $r$  é o grau de crença ( $> 0$ ) por atualmente realizar  $\varphi$  por meio da execução do plano  $\alpha$ .
3.  $best\ plan(\varphi, \alpha, P, A, c_\alpha, r)$  similar ao anterior, mas apenas uma instância com o melhor plano é gerado.

Todo plano precisa ser factível. Este contexto interage com os contextos de intenções, comunicação, desejos e crenças.

O conhecimento  $\varphi$  que o agente tem sobre o estado do mundo e sobre ações que modificam o mundo, são transmitidas do contexto de crenças ao contexto de planejamento como fórmulas de primeira ordem ( $\lceil B\varphi \rceil$ ) com a regra de ponte 3.1.

$$\frac{B : B\varphi}{P : \lceil B\varphi \rceil} \quad (3.1)$$

A regra de ponte 3.2 permite que o contexto de planejamento crie planos. Para isto, ela utiliza desejos positivos e negativos e crenças. O operador  $\nabla$  atribui valor 1 se  $(D^+\varphi) > 0$  e 0 caso contrário. Planos são gerados a partir de ações, visando realizar desejos positivos, e evitar desejos negativos.

$$\frac{\begin{array}{l} D : \nabla(D^+\varphi), \\ D : (D^-\psi, \text{threshold}), \\ P : \text{action}(\alpha, P, A, c), \\ B : (B([\alpha]\varphi), r), \\ B : B(A \rightarrow \neg\psi) \end{array}}{P : \text{plan}(\varphi, \alpha, P, A, c, r)} \quad (3.2)$$

A regra de ponte 3.3 infere a relação entre benefícios e custos de intenções. Diferentes funções podem modelar diferentes comportamentos de indivíduos.

$$\frac{\begin{array}{l} D : (D^+\varphi, d), \\ P : \text{plan}(\varphi, \alpha, P, A, c, r) \end{array}}{I : (I\varphi, f(d, c, r))} \quad (3.3)$$

Regras de ponte podem ser usadas para determinar o relacionamento entre as atitudes mentais e o comportamento atual do agente. Por exemplo, no caso do realismo forte, o conjunto de intenções é um subconjunto do conjunto de desejos, que por sua vez é um subconjunto das crenças. As regras de ponte 3.4 e 3.5 modelam este relacionamento.

$$\frac{B : \neg B\psi}{D : \neg D\psi} \quad (3.4)$$

$$\frac{D : \neg D\psi}{I : \neg I\psi} \quad (3.5)$$

As regras de ponte 3.6 e 3.7 são utilizadas para estabelecer as interações do agente com o ambiente.

$$\frac{I : (I\varphi, i_{max}),}{P : \text{best plan}(\varphi, \alpha, P, A, c\alpha, r)} \quad C : C(\text{does}(\alpha)) \quad (3.6)$$

$$\frac{C : \beta}{B : B\beta} \quad (3.7)$$

### 3.2 BDI + REPAGE

A proposta do trabalho de Pinyol e Sabater-Mir é integrar o modelo de reputação Repage, apresentado na seção 2.4.1.2, em um agente BDI por meio da definição do agente como um sistema multi-contexto.

No contexto de crenças (BC), é utilizada uma lógica dinâmica probabilística, com um conjunto de operadores modais. Em especial, as crenças do agente são diferenciadas das crenças comunicadas por outros agentes. Considerando o aspecto dinâmico, seja  $\alpha$  uma ação, e  $\varphi \in BC$ , fórmulas como  $[\alpha]\varphi$  indicam que após a execução de  $\alpha$ , a fórmula  $\varphi$  é válida. Este contexto troca informações com os contextos Repage, de planejamento e de desejos.

Desejos são explicitamente representados em seu contexto (DC) e refletem os objetivos gerais do agente. A linguagem deste contexto sofre grande inspiração no trabalho apresentado na seção 3.1. A diferença está na explicitação de quem é o desejo, e na separação entre desejos genéricos e concretos. Desejos genéricos definem as preferências gerais do agente. Possuem fórmulas do tipo  $D_i^*\phi$ , onde  $*$  pode ser  $+$  ou  $-$  e  $\phi$  não contém nenhuma ação. Desejos concretos possuem fórmulas do tipo  $D_i^*[\alpha]\phi$ .  $\phi$  é satisfeito com a execução da ação  $\alpha$ . Desejos concretos são gerados por meio de desejos genéricos. Este contexto interage com crenças, planejamento e intenções.

O contexto de intenções (IC) contém representações explícitas das intenções do agente. Elas são geradas de crenças e desejos. Utiliza a mesma linguagem do trabalho *graded BDI* descrito na seção 3.1. Este contexto interage com o contexto de comunicação, desejos e planejamento.

A interface entre o agente e o ambiente é feita utilizando o contexto de comunicação (CC). É utilizada lógica de primeira ordem restrita à cláusulas de Horn, com os predicados *does* para executar ações, e *rec<sub>ij</sub>* para informar que o agente  $i$  recebeu uma comunicação do agente  $j$ . O contexto interage com o ambiente, com os contextos Repage e de intenções.

O contexto de planejamento (PC) é implementado em lógica de primeira ordem restrita a cláusulas de Horn. Detém o predicado *action*, o qual define uma ação primitiva junto com suas condições. O contexto interage com intenções, desejos e crenças.



É definido um contexto chamado Repage (RC), o qual contém a implementação do modelo de confiança de mesmo nome. Os predicados do Repage são especificados em termos de lógica de primeira ordem restritos a cláusulas de Horn. São definidos predicados especiais para imagem (*img*) e reputação (*rep*). Imagem é representada da forma  $img_i(j, r, [V_{w_1}, V_{w_2}, \dots])$ , onde o agente  $i$  possui uma imagem do agente  $j$ , desempenhando o papel  $r$  e avaliando ele de acordo com a distribuição de probabilidade sobre os rótulos  $[V_{w_1}, V_{w_2}, \dots]$ . O modelo BDI+Repage considera 5 rótulos para esta distribuição: *MuitoRuim*, *Ruim*, *Neutro*, *Bom*, *MuitoBom*. O predicado da reputação é o mesmo da imagem, porém representando o que é dito pelo grupo. Este contexto se comunica com as crenças e com o contexto de comunicação.

As regras de ponte 3.8 e 3.9 são responsáveis por gerar as crenças de imagem e reputação. Ambas utilizam o operador de crenças  $B_i$ , representando o que é acreditado pelo agente  $i$ . A segunda utiliza o operador  $S$ , o qual expressa o que é dito pelos agentes do grupo.  $\mathcal{R}_r(j)$  representa a ação executada por  $j$ ,  $\mathcal{T}_{r,w_i}$  é o rótulo sendo considerado e  $V_{w_i}$  é a avaliação do rótulo.

$$\frac{RC : img_i(j, r, [V_{w_1}, V_{w_2}, \dots])}{BC : (B_i([\mathcal{R}_r(j)] \mathcal{T}_{r,w_1}, V_{w_1}))} \quad (3.8)$$

$$BC : (B_i([\mathcal{R}_r(j)] \mathcal{T}_{r,w_2}, V_{w_2}))$$

...

$$\frac{RC : rep_i(j, r, [V_{w_1}, V_{w_2}, \dots])}{BC : (B_i(S([\mathcal{R}_r(j)] \mathcal{T}_{r,w_1}, V_{w_1})))} \quad (3.9)$$

$$BC : (B_i(S([\mathcal{R}_r(j)] \mathcal{T}_{r,w_2}, V_{w_2})))$$

...

As regras 3.10 e 3.11 transformam desejos genéricos em desejos concretos e realistas. Isto é feito com a fusão de desejos genéricos de DC com informação contida em BC, a qual inclui a probabilidade de realizar o desejo por meio da execução de certa ação. O resultado é um desejo cuja graduação mudou, tornando-se mais realística.

$$\frac{DC : (D_i^+ \varphi, d_\varphi)}{BC : (B_i([\alpha] \varphi), p_\psi), BC : B_i(\psi \rightarrow \varphi)} \quad (3.10)$$

$$DC : (D^+ [\alpha] \varphi, g(d_\varphi, p_\psi))$$

$$\frac{DC : (D_i^- \varphi, d_\varphi)}{BC : (B([\alpha] \psi), p_\psi), BC : B_i(\psi \rightarrow \varphi)} \quad (3.11)$$

$$DC : (D^- [\alpha] \varphi, g(d_\varphi, p_\psi))$$

A regra de ponte 3.12 gera intenções. Ela considera tanto o nível de satisfação quanto o custo da ação  $\alpha$ . Uma ação executada para satisfazer certa

fórmula pode causar a realização de desejos negativos. Com isso, esta regra de ponte considera os possíveis desejos negativos que podem ser realizados pela execução da ação. Para cada desejo positivo realista, todos os desejos negativos que podem ser satisfeitos na ação são incluídos.  $\delta$  representa o valor da preferência ou desgosto na realização do desejo. A “força” da intenção que é criada é definida por uma função  $f$ , a qual pode modelar diferentes comportamentos na escolha de intenções.

$$\frac{DC : (D_i^+[\alpha]\varphi, \delta), PC : action(\alpha, P), PC : P}{DC : (D_i^-[\alpha]\psi_1, \delta_{\psi_1}), \dots, (D_i^-[\alpha]\psi_n, \delta_{\psi_n})} \quad (3.12)$$

$$\delta - \sum_{k=1}^n \delta_{\psi_k} \geq 0$$

$$IC : (I_i[\alpha]\varphi, f(\delta, \sum_{k=1}^n \delta_{\psi_k}))$$

A regra de ponte 3.13 instancia uma única intenção, aquela com o grau máximo, e gera a ação correspondente no contexto de comunicação. As regras de ponte 3.14 e 3.15 permitem que os contextos de planejamento (PC) e Repage (RC) conheçam as crenças do agente. Estas informações são importantes para o contexto de planejamento construir planos, ações e suas precondições. No contexto Repage estas fórmulas são utilizadas para configurar os mapeamentos  $\mathcal{R}$  e  $\mathcal{T}$ , e para resolver dissonâncias cognitivas. A regra de ponte 3.16 reflete a transferência de mensagens de imagem e reputação do contexto de comunicação para o contexto repage.

$$\frac{IC : (I_i[\alpha]\varphi, \epsilon_{max})}{CC : does(\alpha)} \quad (3.13)$$

$$\frac{BC : B_i\varphi}{PC : \lceil B_i\varphi \rceil} \quad (3.14)$$

$$\frac{BC : B_i\varphi}{RC : \lceil B_i\varphi \rceil} \quad (3.15)$$

$$\frac{CC : rec_{ij}\varphi}{RC : \lceil rec_{ij}\varphi \rceil} \quad (3.16)$$

### 3.3 ADAPTRUST

Koster, Schorlemmer e Sabater-Mir (2012) propõem um agente como sistema multi-contexto chamado de adapTrust. De acordo com o trabalho, modelos de confiança e reputação são descritos de forma declarativa, de modo a possibilitar que o agente raciocine sobre o modelo. Para isto ser possível, os autores assumem duas propriedades em modelos de confiança e reputação:

(i) natureza funcional; e (ii) existência de pelo menos um parâmetro que ao ser alterado, altera de forma previsível o resultado. Três modelos de reputação são adicionados ao modelo proposto: BRS (COMMERCE; JØSANG; ISMAIL, 2002), ReGreT (SABATER-MIR, 2002) e ForTrust (HERZIG et al., 2010), os autores afirmam que outros podem ser incorporados. Enquanto que os trabalhos descritos nas seções 3.1 e 3.2 tinham por base linguagens proposicionais, este trabalho parte de linguagens de primeira ordem.

O contexto de crenças (BC) utiliza lógica dinâmica de primeira ordem (LDPO). Esta estende a lógica de primeira ordem com modalidade de ações. A semântica de  $[\alpha]\varphi$  é que após a execução da ação  $\alpha$ ,  $\varphi$  é válida. Este contexto interage com os contextos de planejamento, com as regras de prioridade, intenções, sistema de prioridade e interação.

O contexto de desejos (DC) é implementado em lógica de primeira ordem. Desejos são definidos na inicialização do agente e não mudam. Diferentemente dos trabalhos descritos nas seções 3.1 e 3.2, desejos não possuem graus para preferência/rejeição. Ele interage com os contextos de planejamento, intenções e interação.

O contexto de intenções (IC) é definido a partir de uma linguagem LDPO. Ele é formado por fórmulas do tipo  $[\alpha]\varphi$ , representando que o agente tem a intenção de realizar o desejo  $\varphi$  com o plano  $\alpha$ . Este contexto interage com os contextos de regra de prioridade, sistema de prioridade, crenças, desejos e planejamento.

O contexto de interação (CC) representa a interface entre o agente e o ambiente. Ele controla os sensores do agente, executa ações básicas e recebe e envia mensagens. São definidos os predicados especiais  $do(\alpha')$  para representar a ação  $\alpha$  executada pelo agente, e  $sense(\varphi')$  significando que a observação do ambiente  $\varphi$ , em lógica de primeira ordem, foi feita pelo agente. Este contexto interage com os contextos de intenções, desejos, crenças, interação e planejamento, além de interagir com o ambiente.

O contexto de planejamento (PC) é responsável pela busca de planos para alcançar os desejos do agente. O agente assume que planos não falham. Utiliza lógica de primeira ordem restrita a cláusulas de Horn. Três predicados especiais são definidos para representar planos:

1.  $basic\ action(\alpha', \phi'_{Pre}, \phi'_{Post})$  onde  $\alpha$  é uma ação básica, e  $\phi_{Pre}$  e  $\phi_{Post}$  são subconjuntos da linguagem de crenças;
2.  $plan(\alpha', \phi'_{Pre}, \phi'_{Post})$  onde  $\alpha$  é qualquer ação,  $\phi_{Pre}$  e  $\phi_{Post}$  são subconjuntos da linguagem de crenças.
3.  $best\ plan(\psi', \alpha', \phi'_{Pre}, \phi'_{Post})$  correspondendo à noção do agente de melhor instância de um plano para alcançar um desejo  $\psi$ .

Os predicados  $belief(' \phi')$  e  $desire(' \psi')$ , são definidos a partir de crenças e desejos visando escolher planos de acordo com a participação de outros agentes, e assim, permitindo o raciocínio sobre confiança. Este contexto interage com desejos, intenções, comunicação e crenças.

Interação e planejamento utilizam um operador que transforma fbfs-LP (fórmulas bem formadas — Lógica de prioridade), fbfs-LDPO, e conjuntos de fbfs-LDPO em termos de primeira ordem, ou seja, para possibilitar a integração de crenças, desejos e intenções com a interação e com o planejamento.

No contexto de sistema de prioridade (PSC) é armazenado um conjunto de prioridades para cada parâmetro do modelo de confiança e reputação. Por exemplo, ao avaliar um curso, a qualidade do material pode ser mais importante que o custo dele. As prioridades do agente são dependentes dos objetivos do agente. Este contexto interage com crenças, intenções e regras de prioridade.

O contexto de regras de prioridade (PRC) possibilita que o agente adapte suas prioridades. Ele é restrito a cláusulas de Horn, com três predicados especiais:  $belief\_rule$ ,  $role\_rule$  e  $goal\_rule$  para especificar que um conjunto de crenças, um papel ou um objetivo suportam uma ordenação de prioridades. Outros três predicados são usados  $belief$ ,  $role$ , e  $goal$  para representar as atitudes mentais do agente no PRC. Este contexto interage com os contextos de crenças, intenções e sistema de prioridade.

As regras de ponte 3.17 e 3.18 adicionam, respectivamente, uma crença  $\phi$  e um desejo  $\psi$  no contexto de planejamento. Com isto, crenças podem ser utilizadas na busca por planos visando realizar desejos.

$$\frac{BC : \phi}{PC : belief(' \phi')} \quad (3.17)$$

$$\frac{DC : \psi}{PC : desire(' \psi')} \quad (3.18)$$

O contexto de planejamento deduz fórmulas para os desejos com o predicado  $best\ plan$ . Dado um desejo  $\psi$  que o agente deseja realizar e uma base de crenças  $\phi$  em que as precondições  $\phi_{Pre}$  estejam contidas em  $\phi$ . Após a execução de um plano  $\alpha$ , as pós-condições  $\phi_{Post}$  são geradas. A regra de ponte 3.19 é utilizada para mapear este conhecimento para intenções. A regra de ponte 3.20 permite a execução de ações.

$$\frac{PC : best\ plan(' \psi', ' \alpha', ' \phi'_{Pre}, ' \phi'_{Post})}{DC : \psi} \quad (3.19)$$

$$\frac{}{IC : [\alpha]\psi}$$

$$\frac{DC : \psi}{IC : [\alpha]\psi} \quad (3.20)$$

$$CC : do(' \alpha')$$

Após a execução, a regra de ponte 3.21 é usada para que os efeitos da ação do agente sejam adicionadas novamente na base de crenças. A regra de ponte 3.22 é responsável pelo recebimento das entradas dos sensores.

$$\frac{PC : basic\ action(' \alpha', ' \phi'_{Pre}, ' \phi'_{Post})}{CC : do(' \alpha')} \quad (3.21)$$

$$BC : \phi_{Post}$$

$$\frac{CC : sense(' \phi')}{BC : \phi} \quad (3.22)$$

Estas regras de ponte não possuem considerações temporais e assumem que não existe possibilidade de falha. De acordo com os autores, é apresentada uma especificação mínima visando permitir o raciocínio sobre confiança.

Os predicados *belief*, *role* e *goal* nas regras de ponte 3.23, 3.24 e 3.25 são utilizados para representar as atitudes mentais do agente no PRC. Estas regras de ponte são utilizadas para ligar o PRC aos contextos de crenças e intenções.

$$\frac{BC : \phi}{PRC : belief(' \phi')} \quad (3.23)$$

$$\frac{BC : role(\rho)}{PRC : role(\rho)} \quad (3.24)$$

$$\frac{IC : [\alpha]\psi}{PRC : goal(' [\alpha]\psi')} \quad (3.25)$$

As regras de ponte 3.26 e 3.27 estão relacionadas ao raciocínio sobre confiança. De acordo com a primeira, um conjunto de crenças  $\phi$ , uma intenção  $\gamma$ , onde o plano  $\alpha$  é executado para realizar o desejo  $\psi$  sendo que o papel  $r$  é necessário em ações sociais do plano, e um conjunto  $\Upsilon$  de todas regras de prioridade do agente, são usadas para gerar um sistema de prioridade  $PS_{r,\gamma,\phi}$ . A segunda regra de ponte é utilizada para que o modelo de confiança possa calcular a avaliação de confiança de parceiros e adicioná-la ao contexto de crenças.

$$\begin{array}{l}
 BC : \phi \\
 IC : \gamma = [\alpha] \Psi \langle r \rangle \\
 PRC : \Upsilon \\
 \hline
 PSC : PS_{r,\gamma,\phi}
 \end{array} \tag{3.26}$$

$$\begin{array}{l}
 IC : \gamma = [\alpha] \Psi \langle r \rangle \\
 BC : \phi \\
 PSC : PS_{r,\gamma,phi} \\
 \hline
 BC : \phi_{\gamma,r,a}
 \end{array} \tag{3.27}$$

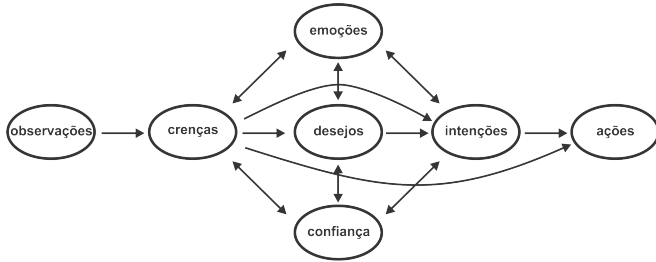
### 3.4 OUTROS TRABALHOS

No trabalho de Bitencourt, Silveira e Marchi (2013) emoções são adicionadas em um modelo de confiança. A proposta dos autores é permear um modelo numérico de confiança com raciocínio simbólico, isto é feito visando tornar a confiança mais dinâmica e próxima do histórico do agente. Os autores não propõem um modelo de confiança, eles adaptam o modelo ReGreT (SABATER-MIR, 2002) para adicionar influência de emoções. Emoções são adicionadas a partir de uma formalização do modelo OCC (ORTON; CLORE; COLLINS, 1990). Além disso, no trabalho é proposta uma função para determinar a intensidade de emoções. O trabalho analisa a integração de emoções em um modelo de confiança, porém a relação entre estes e o raciocínio do agente não é apresentado.

No trabalho de Bosse e Höhle (2011), confiança e emoções são estados mentais adicionados a um agente BDI. O modelo é aplicado ao domínio de futebol virtual (BOSSE; HÖHLE, 2011). A implementação é feita a partir da linguagem de modelagem baseada em agentes LEADSTO (BOSSE et al., 2005). Emoções são geradas seguindo teorias apreciativas, ou seja, a partir da avaliação de eventos relacionados aos objetivos, ou desejos, do agente. Humor é modelado durando mais que emoções e o seu desencadeamento não necessariamente está vinculado a um evento em específico. Tanto emoções quanto humor ficam no intervalo  $[0, 1]$ . Regras são geradas para representar o impacto de emoções em crenças, desejos e intenções. A confiança é baseada em crenças das capacidades do agente. A figura 3 mostra o modelo proposto pelos autores.

Na pesquisa de Bonnefon et al. (2010), os relacionamentos em nível formal de emoções e confiança são apresentados com base em modelos cognitivos. Os autores apresentam um *framework* lógico, formalizando as estruturas cognitivas das emoções de alegria/angústia, esperança e medo, satisfação

Figura 3 – Modelo BDI com Emoções e Confiança.



Fonte: Adaptado de Bosse e Höhle (2011).

e decepção, medo confirmado e alívio. Confiança e desconfiança são baseadas na definição cognitiva apresentada por Castelfranchi e Falcone (2001). No trabalho são apresentadas as relações formais entre confiança e esperança e entre desconfiança e medo. Para testar o modelo, os autores fizeram testes empíricos, os quais constataram que confiança com sucesso implica em satisfação. Porém não confirmaram que confiar sem êxito implica em decepção, e nem que a não confirmação de desconfiança implica em alívio.

### 3.5 ANÁLISE DOS TRABALHOS RELACIONADOS

O trabalho de Casali, Godo e Sierra (2005) descreve um agente BDI, porém não apresenta a integração de confiança e emoções. Entretanto, o formalismo dos autores para crenças, desejos e intenções é usado como base no modelo de agentes desta pesquisa.

Nos trabalhos de Pinyol e Sabater-Mir (2009) e Koster, Schorlemmer e Sabater-Mir (2012), modelos de confiança são integrados em um agente BDI utilizando SMC. O agente proposto neste trabalho além de também integrar confiança em um SMC, adiciona um contexto para representar e manipular emoções.

Bitencourt, Silveira e Marchi (2013), propõem um modelo de confiança e reputação com emoções, enquanto que esta pesquisa analisa o relacionamento destes componentes em um agente BDI.

Na pesquisa de Bosse e Höhle (2011) tanto emoções quanto confiança podem impactar diretamente em crenças, desejos e intenções. No modelo proposto é feita uma análise mais profunda na influência que emoções tem na determinação da confiança.

No trabalho de Bonnefon et al. (2010) é apresentado um *framework* lógico que relaciona emoções e confiança, mais precisamente a influência que confiança tem em emoções. Porém, nesta dissertação é dado ênfase na relação contrária, isto é, na influência de emoções em confiança.

Neste capítulo foram apresentados trabalhos que utilizam SMC em agentes BDI visando realizar o objetivo de definir a estrutura básica do modelo. O modelo de Casali, Godo e Sierra (2005) foi escolhido para tal propósito. Além disto, foram apresentados trabalhos que relacionam confiança com agentes, e confiança com emoções.



## 4 MODELO PROPOSTO

Neste capítulo é apresentada a proposta de um modelo de agentes, baseado em BDI, seguindo a abordagem de sistemas multi-contexto, capaz de integrar confiança e emoções em seu raciocínio. O modelo proposto parte do agente como o SMC desenvolvido por Casali, Godo e Sierra (2005), introduzido na seção 3.1. A escolha por ele está na fácil integração de contextos para emoções e confiança. No contexto de emoções é inserido o modelo WASABI, apresentado na seção 2.5.1. E no contexto de confiança são descritas diretrizes sobre como definir em quem confiar, além disto, é implementado o modelo de confiança e reputação ReGreT descrito na seção 2.4.1.3. Crenças são graduadas com o grau de certeza que o agente tem sobre ela. Desejos são graduados diferenciando os níveis de preferência e/ou rejeição. Intenções são graduadas com a relação custo/benefício em realizar um objetivo do agente.

Crenças especiais chamadas de *juílgamentos* são utilizadas para intermediar a relação entre emoções e confiança. Schwarz e Clore (1988) argumentam que quando pessoas fazem juílgamentos avaliativos, elas inconscientemente perguntam a si mesmas “como eu me sinto sobre (o juílgamento)?”. A partir disto, o modelo proposto, ao gerar um juílgamento sobre uma entidade, vincula o estado emocional atual.

O agente proposto é utilizado, principalmente, em situações onde a confiança não pode ser determinada de forma puramente racional. A ideia é que, em situações onde não existam informações suficiente para determinar a confiança, o agente utilize seu contexto emocional para escolher ou refinar, a escolha por um parceiro. Por exemplo, assumindo que o conjunto de juílgamentos que o agente *tutor1* tem sobre repositórios não é suficiente para definir, de forma puramente racional, qual é o melhor repositório para interagir visando realizar seu objetivo de *aprenderAutômatos*, o histórico emocional representado em crenças pode fazer parte do raciocínio.

**Definição 4.1** (Agente Proposto). O agente multi-contexto é definido como

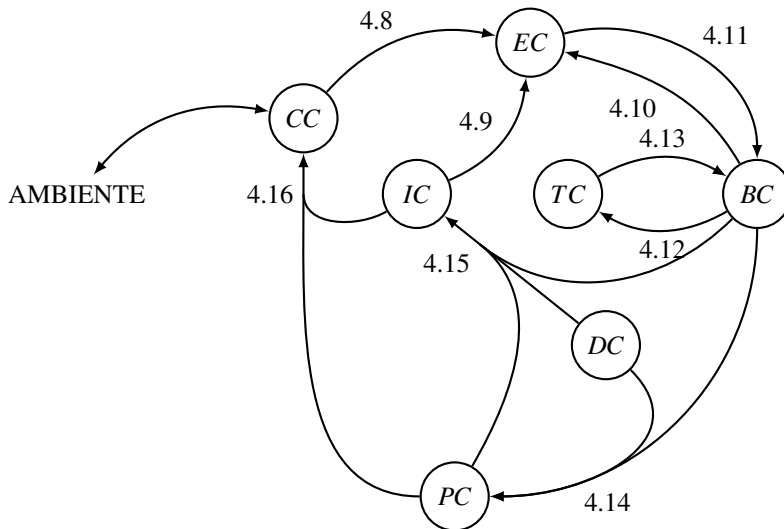
$$AG = \langle \{BC, DC, IC, PC, TC, EC, CC\}, \Delta_{rp} \rangle, \quad (4.1)$$

onde *BC*, *DC*, *IC*, *PC*, *TC*, *EC*, *CC* são, respectivamente, os contextos de crenças, desejos, intenções, planejamento, confiança, emoções e comunicação; e  $\Delta_{rp}$  são as regras de ponte utilizadas para trocar informações entre contextos.

A figura 4 representa o modelo proposto com seus contextos e regras de ponte. O foco deste trabalho está na relação entre os contextos de emoções, crenças e confiança. Desta forma, estes contextos são apresentados com mais

detalhes.

Figura 4 – Modelo de agente proposto. Cada aresta (ou grupo de arestas) representa uma regra de ponte; a numeração refere-se a regra correspondente listada na seção 4.8.



#### 4.1 CONTEXTO DE COMUNICAÇÃO (CC)

Este contexto funciona como uma interface entre os demais componentes internos do agente e o ambiente. Ele é responsável pelas percepções e ações do agente. Assim como nos trabalhos de Casali, Godo e Sierra (2005), Pinyol e Sabater-Mir (2009) e Casali, Godo e Sierra (2011) é utilizada a lógica de primeira ordem restrita a cláusulas de Horn. As percepções do ambiente são divididas em mensagens e observações, sendo representadas, respectivamente, pelos predicados *receive message* e *observe*. As ações executadas são representadas pelo predicado especial *does*. Estes três predicados são utilizados pelo contexto de emoções para gerar estímulos emocionais.

O agente proposto pode receber mensagens sobre experiências de entidades do ambiente com a entidade a ser avaliada. Estas mensagens são julgamentos feitos por terceiros. É assumido que nelas estão contidas as seguintes informações: quem avaliou, quem foi avaliado, qual o contexto, o

que foi avaliado no contexto, a data e o valor da avaliação. Como emoções são algo próprio de cada agente, é assumido que elas não estão presentes em julgamentos comunicados.

## 4.2 CONTEXTO DE EMOÇÕES (EC)

Este contexto representa e manipula as emoções do agente. A literatura da área de emoções em agentes apresenta diversos modelos. Desta forma, ao invés de propor um novo, é utilizado um modelo existente. Dentre os estudados, foi escolhido o WASABI (BECKER-ASANO, 2008) apresentado na seção 2.5.1 para gerenciar o estado emocional do agente. A escolha pelo modelo se deve à flexibilidade de integração com *frameworks* de agentes cognitivos (BECKER-ASANO, 2014), por separar emoções primárias e secundárias, permitindo que tanto ações reativas quanto cognitivas com influência emocional sejam tomadas e por representar emoções de forma que a integração com confiança não precise de manipulações complexas.

O objetivo em integrar emoções no agente, está em permitir que as experiências afetivas do agente façam parte do fator subjetivo na definição de confiança. Embora o impacto direto delas esteja nas crenças do agente, o fator emocional vinculado a estas crenças é adicionado aos algoritmos para cálculo de confiança. Em agentes E-BDI, emoções podem influenciar em vários componentes do raciocínio, porém neste trabalho, apenas o aspecto descrito acima é explorado.

No momento em que uma crença do tipo julgamento é criada, o estado emocional atual do agente é integrado a esta crença. Por exemplo, após o agente *tutor1* observar que o agente *repositório1* estava offline, o estado emocional gera um impulso negativo resultando nas emoções *irritado*(0.175) e *concentrado*(0.19). Estas são armazenadas como parte do julgamento gerado.

Seguindo o modelo WASABI, o contexto de emoções é organizado em dois componentes: *dinâmica emocional* — estímulos positivos e negativos são utilizados para definir emoções, humor e personalidade; e *espaço PAD* — o resultado da dinâmica emocional é mapeado para este componente que avalia quais emoções foram envolvidas com o estímulo, além das suas intensidades. Embora o modelo original considere emoções secundárias, elas não são analisadas nesta pesquisa.

## 4.2.1 Atualização do Estado Emocional

Estímulos emocionais são provenientes dos contextos de comunicação, crenças e intenções. Do contexto de comunicação chegam observações, mensagens e ações executadas. Da perspectiva do contexto de emoções, observações sobre uma entidade do ambiente podem resultar em impulsos emocionais, que são integrados nas crenças de julgamento do avaliado. Mensagens podem representar julgamentos comunicados, e serem impulsos positivos se forem de acordo com o conhecimento do agente, e negativos caso contrário. As ações que o agente executou geram impulsos emocionais, elas provocam alterações nas crenças do agente, e esta alteração gera apreciação para o contexto de emoções. Intenções estão fortemente ligadas aos desejos do agente, elas geram impulsos a partir de sua graduação. O valor de estímulos emocionais provocados por cada fonte é definido de acordo com o ambiente em que o agente está inserido.

### 4.2.1.1 Dinâmica Emocional

A dinâmica emocional descreve a interação entre os fenômenos afetivos de emoções e humor. O primeiro passo da implementação é um arranjo ortogonal destes dois componentes, onde a valência das emoções corresponde ao eixo  $x$ , e a valência do humor ao eixo  $y$ . Seguindo o trabalho de Becker-Asano (2008), é definido um *ponto de referência* na origem, com massa

$$\text{massa}(5000)$$

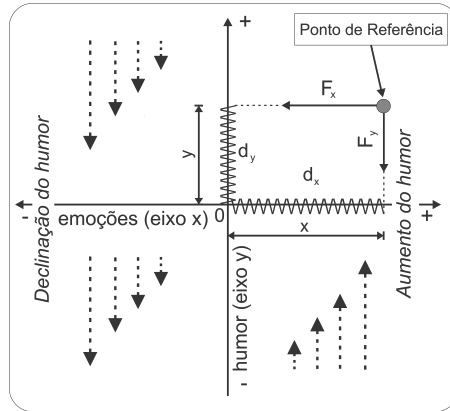
e, enquanto ele for deslocado dela, duas forças independentes são criadas  $F_x$  e  $F_y$ , para trazer seu eixo de volta a origem. Quanto maior for a distância entre o ponto de referência e o eixo de emoções ou humor, maior é a valência do componente. A figura 5 apresenta um exemplo do relacionamento descrito entre emoções e humor.

Os valores constantes para as forças, segundo Becker-Asano (2008) são

$$\begin{aligned} & \text{força\_retorno\_x}(50) \\ & \text{força\_retorno\_y}(10). \end{aligned}$$

Estes valores são importante na estabilização do estado emocional do agente. A força aplicada às emoções é maior, em razão delas durarem menos que o humor.

Figura 5 – Dinâmica emocional do modelo WASABI.



Fonte: Adaptado de Becker-Asano (2008).

Emoções e humor são atualizados, em cada ciclo, separadamente. As equações 4.2 (aceleração), 4.3 (velocidade) e 4.4 (deslocamento) são utilizadas para definir a nova posição de cada eixo. Emoções influenciam no humor do agente, assim, a nova posição em  $y$  utiliza também a equação 4.5, onde  $\alpha$  é um aspecto relacionado a personalidade e quanto maior o seu valor, mais temperamental o agente, quanto menor o valor, mais apático.  $\Delta x$  e  $\Delta y$  representam, respectivamente, a variação em emoções e a variação no humor.

$$a = \frac{F}{m} \quad (4.2)$$

$$v = a * t + ultimaVelocidade \quad (4.3)$$

$$S = v * \Delta t + ultimaPosicao \quad (4.4)$$

$$\frac{\Delta y}{\Delta x} = \alpha . x \quad (4.5)$$

Além da dinâmica emocional, Becker-Asano (2008) adicionam o conceito de *tédio*. Tédio é representado como um terceiro eixo,  $z$ . Ele emerge na ausência de estímulos.

#### 4.2.1.2 Mapeamento para o Espaço PAD

O componente da dinâmica emocional é capaz de fornecer, a qualquer tempo  $t$ , uma 3-upla da forma

$$D(t) = (x_t, y_t, z_t), \text{ com } x_t = [-1, 1], y_t = [-1, 1], z_t = [-1, 0]$$

onde  $x_t$  representa a valência emocional,  $y_t$  a valência de humor, e  $z$  o grau de tédio. De acordo com este aspecto, quanto mais próximo de  $-1$ , mais entediado o agente estará. Estes três valores são utilizados para o mapeamento no espaço PAD. O novo valor de P é

$$P(x_t, y_t) = \frac{1}{2} \cdot (x_t + y_t)$$

A é atualizado para

$$A(x_t, z_t) = |x_t| + z_t$$

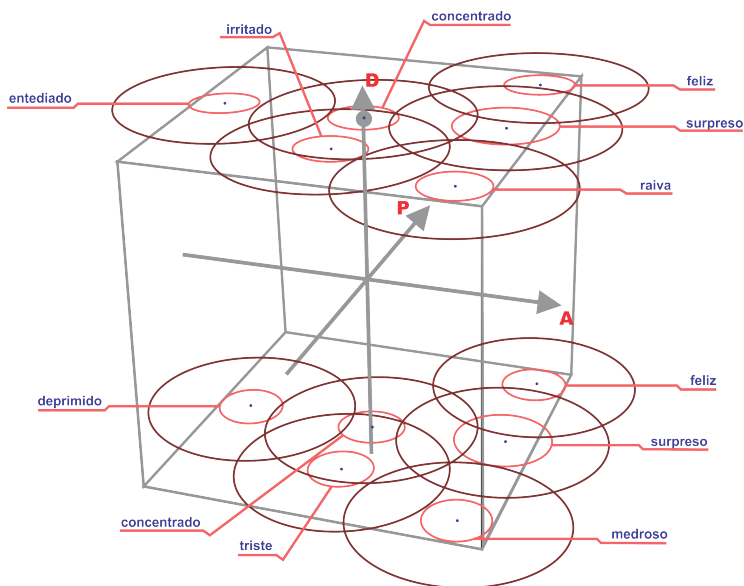
D não é atualizado a partir da dinâmica emocional, ele é um componente do agente, que define se ele está dominante ou submissivo de acordo com circunstâncias externas (BECKER-ASANO, 2008).

Com a definição da posição no espaço PAD, podem ser encontradas as emoções envolvidas. Becker-Asano (2008) apresenta valores padrão para as posições de emoções no espaço. No modelo proposto, estes valores são representados como:

$$\begin{aligned} & raiva(80, 80, 100, 0.75, 0.2, 0.64) \\ & irritado(-50, 0, 100, 0.75, 0.2, 0.64) \\ & entediado(0, -80, 100, 0.75, 0.2, 0.64) \\ & concentrado(0, 0, \pm 100, 0.75, 0.2, 0.64) \\ & deprimido(0, -80, -100, 0.75, 0.2, 0.64) \\ & medroso(-80, 80, 100, 0.25, 0.2, 0.64) \\ & feliz(80, 80, \pm 100, 0.75, 0.2, 0.64) \\ & triste(-50, 0, -100, 0.75, 0.2, 0.64) \\ & surpreso(10, 80, \pm 100, 0.0, 0.3, 0.64) \end{aligned}$$

onde os três primeiros valores de cada predicado representam, respectivamente, aos valores de prazer, ativação e dominância, sendo as coordenadas que definem a localização da emoção. O quarto elemento corresponde ao va-

Figura 6 – Emoções primárias no espaço PAD.



Fonte: Adaptado de Becker-Asano (2014).

lor de sua intensidade máxima. O quinto e o sexto elementos representam os raios de saturação e ativação. Os três últimos valores podem ser alterados visando aumentar/diminuir a apreciação de uma emoção. A figura 6 apresenta as emoções no espaço PAD bem como suas regiões de saturação e ativação.

Quando o ponto de referência estiver dentro do círculo menor de uma emoção, a intensidade desta é assumida como o seu valor de intensidade máxima. Quando o ponto estiver fora do círculo menor, mas dentro do círculo maior, a intensidade desta emoção é calculada a partir da função 2.17. Por exemplo, na figura 6, o ponto de referência está na posição PAD  $(0, 0, 100)$ , duas emoções estão envolvidas neste momento: *concentrado* com intensidade máxima  $(0.75)$ , e *irritado* com intensidade de  $0.23$ .

#### 4.3 CONTEXTO DE CRENÇAS (BC)

As crenças representam o conhecimento que o agente tem sobre o mundo. Elas podem possuir graduações representando o grau de certeza de

sua credibilidade. É definido o predicado  $J$  para representar os julgamentos feitos pelo agente sobre interações com outros agentes, instituições ou grupos. Crenças sobre a confiança que o agente tem em entidades do ambiente são definidas com o predicado  $C$ .

### 4.3.1 A Linguagem de Crenças

A linguagem de crenças apresentada por Casali, Godo e Sierra (2005), é estendida para contemplar julgamentos e confiança. Primeiro é apresentada a linguagem original, para então adicionar os componentes necessários para confiança.

Casali, Godo e Sierra (2005) partem de uma linguagem proposicional  $\mathcal{L}$ , definida sobre um conjunto contável de variáveis proposicionais  $VP$  e os conectivos de negação ( $\neg$ ) e implicação ( $\rightarrow$ ). Esta linguagem é estendida para  $\mathcal{L}_D$  visando adicionar modalidades de ações.  $\mathcal{L}_D$  é definida como:

1.  $\Pi_0 \subset \Pi$  (Ações elementares são planos)
2. se  $\alpha, \beta \in \Pi$ , então  $\alpha; \beta \in \Pi$  (A concatenação de ações também é um plano)
3. se  $\alpha, \beta \in \Pi$ , então  $\alpha \cup \beta \in \Pi$  (Disjunção não determinística)
4. se  $\alpha \in \Pi$ , então  $\alpha^* \in \Pi$  (Iteração)
5. se  $A$  é uma fórmula, então  $A? \in \Pi$  (Teste)
6. se  $p \in VP$ , então  $p \in \mathcal{L}_D$
7. se  $\varphi \in \mathcal{L}_D$ , então  $\neg\varphi \in \mathcal{L}_D$
8. se  $\varphi, \psi \in \mathcal{L}_D$ , então  $\varphi \rightarrow \psi \in \mathcal{L}_D$
9. se  $\alpha \in \Pi$  e  $\varphi \in \mathcal{L}_D$ , então  $[\alpha]\varphi \in \mathcal{L}_D$

onde  $\Pi_0$  representa um conjunto de ações elementares,  $\Pi$  é o conjunto de planos (ações compostas).  $[\alpha]A$  pode ser interpretado como “após a execução da ação  $\alpha$ ,  $A$  é verdadeiro”.

Esta linguagem crisp é estendida pelos autores com o operador modal fuzzy  $B$ . Se  $\varphi$  é uma proposição em  $\mathcal{L}_D$ ,  $B\varphi$  pode ser lido como “ $\varphi$  é acreditável”. Assim, fórmulas de BC são:

1. *Crisp (não B-modal)*. Se  $\varphi \in \mathcal{L}_D$ , então  $\varphi \in BC$



2. *B-modal*: São construídas a partir de fórmulas modais elementares  $B\phi$ , onde  $\phi$  tem valor crisp, e constantes verdades  $\bar{r}$ , para cada racional  $r \in [0, 1]$ , utilizando lógica infinitamente valorada:

- se  $\phi \in \mathcal{L}_D$ , então  $B\phi \in BC$
- se  $r \in Q \cap [0, 1]$ , então  $\bar{r} \in BC$
- se  $\Phi, \Psi \in BC$ , então  $\Phi \rightarrow_L \Psi \in BC$  e  $\Phi \& \Psi \in BC$  (onde  $\&$  e  $\rightarrow_L$  correspondem, respectivamente, a conjunção e implicação da lógica de Łukasiewicz).

onde  $Q$  representa o conjunto dos números racionais.

Na lógica de Łukasiewicz, uma fórmula  $\phi \rightarrow_L \psi$  é 1-verdade sse o valor verdade de  $\psi$  é maior ou igual ao valor de  $\phi$ . Fórmulas do tipo  $\bar{r} \rightarrow B\phi$  representam que a probabilidade de  $\phi$  é no mínimo  $r$ , a partir de agora, esta notação é  $(B\phi, r)$ .

Neste ponto, a linguagem proposta por Casali, Godo e Sierra é estendida para suportar o modelo proposto. Os predicados  $J$  e  $C$  são adicionados para representar, respectivamente, julgamentos e confiança. O predicado  $J$  tem aridade 7, e é da forma

$$J(X, Y, C_j, P_c, A_r, \{E\}, T),$$

onde  $X$  é o agente que criou o julgamento,  $Y$  é a entidade avaliada,  $C_j$  é o contexto do julgamento,  $P_c$  é a propriedade deste contexto sendo avaliada,  $A_r \in [-1, 1]$  é a avaliação racional sobre esta propriedade,  $\{E\}$  é o conjunto das emoções envolvidas no momento em que o julgamento foi realizado e  $T$  é o tempo em que o julgamento foi criado. Um julgamento pode ser lido como: “ o agente  $X$  avaliou a entidade  $Y$ , considerando a propriedade  $P_c$  do contexto do julgamento  $C_j$ , no tempo  $T$ , fazendo avaliação racional  $A_r$  e sentindo as emoções  $\{E\}$ . O número de julgamentos que o agente armazena em seu contexto de crenças é dependente do domínio. Por exemplo, em um ambiente dinâmico, onde o agente cria julgamentos em curtos períodos de tempo, pode ser muito custoso manter todos eles. Desta forma, o agente precisa de planos que removam julgamentos antigos, ou então julgamentos que não estão relacionados aos seus objetivos.

O predicado  $C$  tem aridade 5, e é da forma

$$C(Y, C_c, Obj, V, T),$$

onde  $Y$  é a entidade confiada,  $C_c$  é o contexto da confiança,  $Obj$  é o objetivo dependente de  $Y$ ,  $V \in [-1, 1]$  é o valor que representa quanto o agente confia

em  $Y$ , e  $T$  é o tempo em que a confiança foi definida. A crença de confiança pode ser lida como: “No tempo  $T$ , o agente confia na entidade  $Y$  com valor  $V$  considerando o contexto  $C_c$ , visando realizar o objetivo  $Obj$ .” Cabe ao agente decidir o número de novos julgamentos necessários para atualizar sua crença  $C$ .

Esta definição de confiança é bem próxima a de Castelfranchi e Falcone (2010), onde confiança é uma relação de, pelo menos, cinco partes: (i) um agente que confia  $X$ ; (ii) uma entidade confiada  $Y$ ; (iii) uma ação  $\alpha$  executada por  $Y$  e seu resultado  $p$ ; (iv) um objetivo  $g$ ; e (v) um contexto  $C$ . Estes fatores podem ser resumidos como: o agente  $X$  confia na entidade  $Y$  no contexto  $C$  para executar a ação  $\alpha$  e obter o resultado  $p$ , visando realizar, pelo menos parcialmente, o objetivo  $g$ .

#### 4.4 CONTEXTO DE CONFIANÇA (TC)

Este contexto determina a confiança do agente em outras entidades. Isto é feito com a utilização de um modelo de confiança, de informações do próprio contexto, além de informações fornecidas pelo contexto de crenças. O modelo de confiança utilizado é uma adaptação do ReGreT, descrito na subseção 2.4.1.3. As informações que estão em TC podem ser vieses, normas, descrições de domínios, etc. O contexto de crenças fornece julgamentos do próprio agente ou de terceiros, e podem sofrer influência emocional.

No trabalho de Hoelz e Ralha (2014) os autores criam um modelo para confiança, ele é definido como

$$m_{conf} = \{id, v, n, dc, ac, td\}$$

representando, respectivamente, interação direta, uso de vieses, normas, dimensões e contextos, avaliação de confiança e tomada de decisão. Com base nisto, as informações utilizadas no contexto de confiança são:

1. *Interação direta*: o resultado de uma interação entre o agente com outra entidade do ambiente. Este resultado gera um ou mais julgamentos, que podem sofrer influência emocional.
2. *Observação direta*: observação da entidade alvo, avaliando seu comportamento. Também gera um ou mais julgamentos com influência emocional.
3. *Informação de testemunho*: julgamento feito por outra entidade do ambiente da entidade alvo. Esta não sofre influência emocional, porém

ao receber a informação, o agente pode gerar um julgamento com influência emocional sobre a entidade que forneceu a informação. Possui custo alto por precisar interagir com o ambiente (HOELZ; RALHA, 2014).

4. *Viés*: uma preferência, positiva ou negativa, interna do agente. Por exemplo, um viés emocional pode permitir que o agente dê mais atenção aos julgamentos negativos do que positivos (SMITH et al., 2006).
5. *Norma*: descrição de obrigações, permissões e proibições que influenciam o comportamento de um agente dentro de uma sociedade (KOLLINGBAUM, 2005).
6. *Dimensões e Contexto*: dimensões podem ser combinadas para gerar uma confiança mais global. Por exemplo, as confianças sobre didática e conhecimento sobre autômatos podem ser combinadas para gerar uma confiança em relação a ser professor de autômatos. Contextos descrevem com maior precisão um cenário envolvido na definição de confiança.
7. *Avaliação da confiança*: na presença de mais de um modelo de confiança, o agente pode fazer uso das informações que possui para decidir qual modelo utilizar. Neste trabalho foi usado o ReGreT, desta forma, a avaliação fica restrita a manipulação dos parâmetros do ReGreT.
8. *Resultado na tomada de decisão*: criadas as crenças de confiança, cabe ao contexto de planejamento e de intenções decidir com quem interagir.

Os itens 1, 2, 3 ficam armazenados no contexto de crenças. Eles são obtidos por meio de interações com o meio externo. Os itens 4, 5, 6 são representados no contexto de confiança em lógica descritiva. A escolha por lógica descritiva está na clara visão destes itens como domínios. Além disso, a descrição em lógica descritiva permite que, apesar de serem do agente, nada impede que sejam definidas a partir de descrições compartilhadas.

O item 7 é feito por meio do modelo ReGreT. Este foi escolhido por ser um modelo numérico, além disso, ele permite que emoções influenciem na confiança de forma baseada e não baseada em crenças. A influência baseada em crenças, de acordo com o modelo proposto, é feita por meio da apreciação afetiva de julgamentos na definição da reputação individual. A influência não baseada em crenças pode ser feita dinamicamente nas equações 2.13 e 2.14. Este aspecto foge do escopo do trabalho. O item 8 não é feito no contexto de confiança e é a etapa final do processo.

#### 4.4.1 Adicionando o Modelo ReGreT

Como descrito na seção 2.4.1.3, ReGreT é um modelo numérico, que organiza confiança e reputação em três dimensões: individual, social e ontológica. Visando integrar emoções, a dimensão individual foi alterada. A partir de agora, o conceito de reputação individual definido em Sabater e Sierra (2001) e apresentado na seção 2.4.1.3, é chamado de *confiança direta*. Esta nomenclatura é utilizada pelo próprio autor em sua tese (SABATER-MIR, 2002).

A confiança direta (CD) é calculada a partir dos julgamentos feitos pelo próprio agente, e com isto, nesta pesquisa, é onde emoções influenciam. No momento em que o agente computa o valor da confiança direta, ele possui tanto avaliações racionais quanto as apreciações emocionais. Isto possibilita que estes dois componentes sejam combinados de várias formas. A seguir será descrita uma possível combinação.

Um julgamento qualquer é emocionalmente positivo ou negativo e composto por uma ou mais emoções. Todas emoções envolvidas nele tem suas intensidades mapeadas para o intervalo  $[0, 1]$  e, a cada uma delas, é atribuído um peso  $p$  no julgamento. A função 4.6 é utilizada para integrar as todas emoções do julgamento em um componente definido como *viés emocional* ( $ve$ ), o qual é utilizado na definição da confiança.

$$ve = \begin{cases} \sum_{E_i \in \text{Emoções}} i_i \cdot p_i & \text{se julgamento positivo} \\ -(\sum_{E_i \in \text{Emoções}} i_i \cdot p_i) & \text{caso contrário} \end{cases} \quad (4.6)$$

onde  $p_1 + \dots + p_n = 1$ , *Emoções* é o conjunto das emoções envolvidas com intensidade mapeada no intervalo  $[0, 1]$ ,  $i_i$  é a intensidade da emoção  $E_i$ ,  $ve \in [-1, 1]$ , representando que, se for um julgamento positivo ele ficará no intervalo  $[0, 1]$  e se for negativo  $[-1, 0]$ .

Por exemplo, em um julgamento negativo, as emoções presentes são *raiva*(0.41) e *irritado*(0.22). No intervalo  $[0, 1]$  elas possuem, respectivamente, os valores 0.54 e 0.29. O agente atribuiu peso maior para a raiva, resultando no valor do viés emocional  $ve = -(0.7 \cdot 0.54 + 0.3 \cdot 0.29) = -0.465$ .

O próximo passo é vincular os julgamentos e seus vieses emocionais ao cálculo da confiança direta. A função 4.7 apresenta esta integração.

$$CD_{a \rightarrow b(\varphi)} = \sum_{J_i \in \text{Julgamentos}} \rho(t, t_i) \cdot ((\alpha \cdot W_i) + ((1 - \alpha) \cdot ve_i)), \quad (4.7)$$

onde *Julgamentos* é o conjunto de julgamentos feitos pelo agente sobre uma entidade em um determinado contexto,  $\rho(t, t_i)$  é a função definida na seção 2.4.1.3, a qual atribui peso maior para tempos mais recentes,  $\alpha \in [0, 1]$  é o peso da influência racional no julgamento, ele pode ser influenciado por vieses e normas,  $W_i \in [-1, 1]$  refere-se à avaliação racional feita no julgamento  $i$  e  $ve_i \in [-1, 1]$  é o viés emocional do julgamento  $i$ .

Os demais componentes na definição de confiança seguem a versão original do modelo ReGreT, descritos na seção 2.4.1.3. O resultado deste cálculo é um valor no intervalo  $[-1, 1]$ , o qual é o valor atribuído na crença de confiança. Dado um contexto, TC fornece a BC uma crença de confiança para cada entidade avaliada. Os contextos de planejamento e de intenções fazem uso destas crenças visando escolher uma entidade e interagir com ela.

## 4.5 CONTEXTO DE DESEJOS (DC)

Os desejos representam os objetivos gerais e, às vezes, controversos do agente. Eles são separados em positivos e negativos. Enquanto que os desejos positivos o agente espera que sejam alcançados, os negativos ele gostaria que não se concretizem. Além disto, o agente pode desejar que um objetivo se concretize mais do que outro, assim, são adicionadas graduações aos desejos.

A linguagem utilizada para definir este contexto é a mesma definida por Casali, Godo e Sierra (2005).

### 4.5.1 A Linguagem de Desejos

Seja  $\mathcal{L}$  uma linguagem proposicional, formada por um conjunto contável  $P$  de variáveis proposicionais, e com os conectivos  $\rightarrow$  e  $\neg$ . Esta linguagem é estendida com os operadores modais fuzzy  $D^+$  e  $D^-$ , onde  $D^+ \phi$  é lido como “ $\phi$  é positivamente desejado” e  $D^- \phi$  é lido como “ $\phi$  é negativamente desejado”, e o grau de verdade representa o nível de satisfação/desgosto em  $\phi$  se tornar verdade. Assim como no contexto de crenças, lógica modal infinitamente valorada de Łukasiewicz é utilizada para formalizar os desejos graduados. Esta linguagem é estendida com o conectivo  $\Delta$ . Representando que para cada modal  $\phi$ , se  $\phi$  tem valor  $< 1$  então  $\Delta\phi$  recebe 0; caso contrário,  $\Delta\phi$  recebe 1. Ou seja,  $\Delta\phi$  é uma fórmula booleana com dois possíveis valores. A linguagem para desejos  $\mathcal{L}_{DC}$  é definida como:

1. *Crisp* (não modal): fórmulas de  $\mathcal{L}$
2. *Infinitamente valorada* (modal): construídas a partir  $D^+ \phi$  e  $D^- \phi$ , onde

$\varphi$  é proveniente de  $\mathcal{L}$  e constantes verdades  $\bar{r}$  para cada racional  $r \in [0, 1]$ :

- Se  $\varphi \in \mathcal{L}$ , então  $D^- \varphi, D^+ \varphi \in DC$
- Se  $r \in Q \cap [0, 1]$ , então  $\bar{r} \in DC$
- Se  $\Phi, \Psi \in DC$ , então  $\Phi \rightarrow_{\perp} \Psi \in DC$  e  $\Phi \& \Psi \in DC$

Da mesma forma que em  $BC$ ,  $(D\psi, \bar{r})$  denota  $\bar{r} \rightarrow_{\perp} D\psi$ .  $(D\psi^+, \alpha)$  e  $(D\phi^-, \beta)$  representam que  $\psi$  é preferido e  $\phi$  é rejeitado, com graus  $\alpha$  e  $\beta$ , respectivamente. No caso de  $(D\psi^+, 1)$ , o agente tem preferência máxima para que  $\psi$  se torne verdade. Analogamente, para o caso de  $(D\phi^-, 1)$ , o agente tem rejeição máxima para que  $\phi$  se torne verdade.

#### 4.6 CONTEXTO DE INTENÇÕES (IC)

Este contexto representa as intenções do agente, as quais são geradas a partir de crenças e desejos. Intenções são graduadas com o intuito de avaliar os custos e benefícios envolvidos em ações. A linguagem utilizada é a definida por Casali et al. (CASALI; GODO; SIERRA, 2005; CASALI et al., 2008).

##### 4.6.1 A Linguagem de Intenções

A linguagem de intenções parte da mesma linguagem  $\mathcal{L}$  utilizada nos contextos de crenças e desejos. Este contexto assume que o agente tem um conjunto finito de ações ou planos  $\Pi_0$  para realizar os desejos. Para cada  $\alpha \in \Pi_0$ , é adicionado o operador modal  $I_\alpha$ , no qual o grau de verdade de uma fórmula  $I_\alpha \varphi$  representa a dedicação para realizar o desejo  $\varphi$  por meio de uma ação particular  $\alpha$ . Outro operador adicionado é  $I$ , no qual o grau de verdade  $I\varphi$  representa o grau da intenção com a qual o agente pretende realizar  $\varphi$  utilizando o melhor plano em  $\Pi_0$ . A linguagem para intenções  $\mathcal{L}_{IC}$  é definida como:

- Se  $\varphi \in \mathcal{L}$  então  $\varphi \in \mathcal{L}_{IC}$
- Se  $\alpha \in \Pi_0$  então  $C_\alpha \in \mathcal{L}_{IC}$
- Se  $\varphi \in Sat(\mathcal{L})$  e  $\alpha \in \Pi$  então  $I_\alpha \varphi, I\varphi \in \mathcal{L}_{IC}$
- Se  $r \in Q \cap [0, 1]$  então  $\bar{r} \in \mathcal{L}_{IC}$

- Se  $\phi, \psi \in \mathcal{L}_{IC}$  então  $\phi \rightarrow_{\mathcal{L}} \psi \in \mathcal{L}_{IC}$  e  $\neg_{\mathcal{L}} \phi \in \mathcal{L}_{IC}$

onde  $C_{\alpha}$  é o custo da ação  $\alpha$ ,  $Sat(\mathcal{L})$  representa o conjunto satisfazível de  $\mathcal{L}$ . Desta forma, o valor de cada intenção depende da sua fórmula que está sendo realizada e dos benefícios e custos relacionados.

#### 4.7 CONTEXTO DE PLANEJAMENTO (PC)

O contexto de planejamento constrói planos para realizar os desejos do agente. A linguagem utilizada é baseada em lógica de primeira ordem restrita a cláusulas de Horn. Casali, Godo e Sierra (2005) definem três predicados especiais: *action*, *plan* e *best plan*. Estes predicados são ajustados para integrar confiança na tomada de decisão. O grau de certeza  $r$  que o agente tem sobre suas crenças, como definido no modelo de Casali, Godo e Sierra (2005), é abstraído visando enfatizar a influência de confiança no planejamento.

O predicado *action*, com aridade 4, é da forma

$$action(\alpha, pre\ condition, post\ condition, c_{\alpha}),$$

onde  $\alpha \in \Pi_0$  é uma ação elementar, *pre condition* é o conjunto das condições necessárias para que a ação  $\alpha$  seja executada; *post condition* são as pós-condições e  $c_{\alpha} \in [-1, 1]$  representa o valor da confiança ao executar a ação  $\alpha$ . O valor de  $c_{\alpha}$  é atualizado dinamicamente pela crença de confiança relacionada ao agente envolvido na ação. Caso não exista tal crença sobre a ação, o valor é definido como 0.

O predicado *plan* tem aridade 5, e é da forma

$$plan(\varphi, \alpha, pre\ condition, post\ condition, c_{\alpha}),$$

onde  $\alpha \in \Pi$  é uma ação composta representando o plano para realizar  $\varphi$ , *pre condition* são as precondições de  $\alpha$ , *post condition* são as pós-condições e  $c_{\alpha}$  é a confiança na ação  $\alpha$ .

O terceiro predicado, *best plan*, é da forma

$$best\ plan(\varphi, \alpha, pre\ condition, post\ condition, c_{\alpha}),$$

onde os parâmetros são os mesmos do predicado *plan*, porém, ele representa a instância de melhor plano.

Planos precisam ser factíveis, ou seja, o estado atual do mundo deve satisfazer as precondições. Um plano deve tornar verdade o desejo positivo, o qual justifica sua criação, e deve evitar desejos negativos.

#### 4.8 REGRAS DE PONTE

A comunicação entre os contextos é feita a partir de regras de ponte. As arestas representadas na figura 4 representam as regras de ponte que são descritas nesta seção. Dois contextos só podem trocar informação se existir uma regra que faça o intermédio. Quando uma fórmula de uma unidade é adicionada em outra, ela é circundada por  $[\cdot]$ . Isto significa que esta fórmula é convertida em termos de primeira ordem. Esta abordagem foi proposta por Giunchiglia e Serafini (1994) para modelar lógica modal como teorias de primeira ordem.

1. A percepção  $\varphi$ , obtida do ambiente por meio de sensores pelo contexto de comunicação (CC), é direcionada para o contexto de emoções (EC), onde gera uma crença que é utilizada para avaliar o componente emocional do agente. A mensagem  $\psi$  recebida, quando for uma avaliação, é adicionada a EC como um julgamento  $J$ .

$$\frac{CC : observe(\varphi)}{EC : (B\varphi, r)}, \frac{CC : receive\ message(\psi)}{EC : J\psi} \quad (4.8)$$

2. A intenção  $I$  que será executada, com o plano ou ação  $\alpha$  para o desejo  $\varphi$ , e custo-benefício  $r$ , é informada pelo contexto de intenções (IC) para EC visando gerar estímulo emocional.

$$\frac{IC : (I_{\alpha}\varphi, r)}{EC : (I_{\alpha}\varphi, r)} \quad (4.9)$$

3. A alteração do conhecimento do agente sobre o mundo pode gerar estímulos emocionais. Por isso, a crença  $\varphi$  com grau  $r$  do contexto de crenças (BC), representando o conhecimento antigo, é enviada a EC visando comparar com a nova percepção.

$$\frac{BC : (B\varphi, r)}{EC : (B\varphi, r)} \quad (4.10)$$

4. As crenças com grau de certeza  $r$ , obtidas por percepções de CC, e comunicadas a EC, são transferidas a BC, onde é feita a revisão de crenças. Embora todas as percepções sejam transferidas de CC para EC e depois de EC para BC, não são todas elas que sofrem influência emocional. Apenas julgamentos sofrem influência emocional, eles são criados em EC e transferidos para BC.



$$\frac{EC : (B\varphi, r)}{BC : (B\varphi, r)}, \frac{EC : J\psi}{BC : J\psi} \quad (4.11)$$

5. A confiança em uma entidade é definida pelo contexto de confiança (TC) a partir de julgamentos que estão em BC. É necessária uma regra de ponte que transfira este conhecimento e converta-o em termos de primeira ordem, visando o cálculo da confiança.

$$\frac{BC : J\varphi}{TC : \text{julgamento}(\lceil J\varphi \rceil)} \quad (4.12)$$

6. A crença  $C$  de confiança criada por TC a partir dos julgamentos é transferida a BC, onde fica armazenada.

$$\frac{TC : C\varphi}{BC : C\varphi} \quad (4.13)$$

7. Planos são criados pelo contexto de planejamento (PC) considerando desejos positivos  $D^+\varphi$  e negativos  $D^-\varphi$  e seus graus  $d$ , crenças  $B\Phi$  com seu grau  $r$  de certeza e quando for um plano sobre interagir com uma entidade, a crença  $C$  de confiança também é utilizada. Esta regra é a definida por Casali, Godo e Sierra (2011), acrescida com a possibilidade de transferir predicados de confiança. Esta regra de ponte também transforma fórmulas modais em termos de primeira ordem, uma vez que PC é implementado nesta lógica.

$$\frac{DC : (D^+\varphi, d)}{PC : \text{pos desire}(\lceil (D^+\varphi, d) \rceil)}$$

$$\frac{DC : (D^-\psi, d)}{PC : \text{neg desire}(\lceil (D^-\psi, d) \rceil)} \quad (4.14)$$

$$\frac{BC : (B\Phi, r)}{PC : \text{belief}(\lceil (B\Phi, r) \rceil)}$$

$$\frac{BC : C\theta}{PC : \text{confianca}(\lceil C\theta \rceil)}$$

8. Planos, desejos e crenças são utilizados para gerar intenções.  $(B[\alpha]\varphi, r)$  significa que após a execução da ação  $\alpha$ ,  $\varphi$  será verdade, com grau  $r$  de certeza. A função  $f(d, r, c_\alpha) = \frac{d+r+c_\alpha}{3}$  é utilizada para avaliar o custo/benefício de uma intenção. Intuitivamente, quanto mais o agente desejar, confiar e acreditar, maior a prioridade da intenção. Esta re-

gra sofre forte influência do trabalho de (CASALI; GODO; SIERRA, 2011). A função  $f$  é alterada para contemplar confiança.

$$\begin{array}{l}
 DC : (D^+ \varphi, d), \\
 BC : (B[\alpha]\varphi, r), \\
 \hline
 PC : \textit{best plan}(\varphi, \alpha, \textit{pre condition}, \textit{post condition}, c_\alpha) \\
 \hline
 IC : (I_\alpha \varphi, f(d, r, c_\alpha))
 \end{array} \quad (4.15)$$

9. Uma vez definida a intenção, ela é comunicada a CC que é o responsável pela ação. Esta regra é definida no trabalho de (CASALI; GODO; SIERRA, 2011).

$$\begin{array}{l}
 PC : \textit{best plan}(\varphi, \alpha, \textit{pre condition}, \textit{post condition}, c_\alpha), \\
 \hline
 IC : (I_\alpha \varphi, i_\varphi), IC : (I\varphi, i_\varphi) \\
 \hline
 CC : \textit{does}(\alpha, I_\alpha)
 \end{array} \quad (4.16)$$

Neste capítulo foi apresentado um modelo de agentes baseado em BDI que integra confiança e emoções. Ele partiu do agente BDI como SMC proposto por Casali, Godo e Sierra (2005). A este foi adicionado um contexto para manipular informações sobre confiança, e com isto realizar o objetivo específico de desenvolver um contexto para reger as informações sobre confiança e integrá-lo aos componentes do agente. Foi incluído um segundo contexto, este para representar e manipular emoções, e assim, alcançar o objetivo referente a integração de emoções. Além disto, foram adicionadas regras de ponte para comunicação entre contextos, realizando, desta forma, o objetivo geral.

## 5 DINÂMICA DO MODELO

Neste capítulo são apresentados dois estudos de caso utilizados para exemplificar a dinâmica do modelo proposto. A avaliação quantitativa do modelo proposto foge do escopo desta dissertação, uma vez que, para isso, é necessária a criação de uma ferramenta que implemente o agente. O modelo de agente foi definido a partir do trabalho de Casali, Godo e Sierra (2005), o qual está servindo de inspiração para uma extensão do *framework* para implementação de agentes Jason (BORDINI; HÜBNER; WOOLDRIDGE, 2007) chamada G-JASON (BIGA; CASALI, 2013). Desta forma, o enfoque da avaliação está no ciclo de raciocínio do agente, na definição de julgamentos com influência emocional e nas crenças de confiança.

### 5.1 PROCURANDO POR UM MECÂNICO

Suponha um cenário onde Bob é um agente e ele acredita que seu carro está quebrado.

$$(B(\neg\text{carro}(\text{funcionando})), 1)$$

Um desejo positivo de Bob é que seu carro sempre esteja funcionando.

$$(D^+(\text{carro}(\text{funcionando})), 0.7)$$

Assumindo que Bob não saiba consertar seu carro, ele precisa delegar esta tarefa para ter seu objetivo realizado. Suponha que Bob possua uma *crença de confiança* sobre seu mecânico chamado Steve. De acordo com esta crença, no tempo 29/06/2015, Bob confia plenamente (1) em Steve como mecânico, de modo que, seu desejo de ter o carro funcionando será satisfeito se ele delegar esta tarefa para Steve.

$$C(\text{steve}, \text{mecânico}, \text{carro}(\text{funcionando}), 1, 29/06/2015)$$

Bob possui uma crença afirmando que após interagir com Steve, ele terá seu carro funcionando com credibilidade 1.

$$(B([\text{interagirComSteve}]\text{carro}(\text{funcionando})), 1)$$

Então, ao se deparar com seu carro quebrado, Bob seleciona seu plano de interagir com Steve.

```

action(
  interagirComSteve, ¬carro(funcionando) ∧ disponível(steve),
  carro(funcionando), 1)
plan(
  carro(funcionando), interagirComSteve,
  ¬carro(funcionando) ∧ disponível(steve), carro(funcionando), 1)

```

Mas Steve está fora da cidade, e não pode consertar o carro de Bob. Este evento dispara em Bob a criação de julgamentos sobre a indisponibilidade de Steve, e também, causa a diminuição da credibilidade na crença relacionada ao carro estar funcionando sempre que interagir com Steve.

```

(B([interagirComSteve]carro(funcionando)), 9)
J(bob, steve, mecânico, disponibilidade, -1,
  {raiva = 0.75}, 30/06/2015)
J(bob, steve, mecânico, qualidadeDoServiço, -1,
  {raiva = 0.75}, 30/06/2015)

```

Neste momento, Bob procura por outro mecânico em seu contexto de crenças, mas, infelizmente ele não tem crenças ou julgamentos sobre outros mecânicos. Então Bob procura por informações sobre mecânicos perguntando para outros agentes, por meio do predicado *does* do contexto de comunicação, e fazendo observações com o predicado *observe*. As mensagens recebidas de outras entidades com julgamentos chegam ao contexto de comunicação por meio do predicado *receive message*, e são enviadas ao contexto de emoções. Estas não são influenciadas pelas emoções do agente, uma vez que elas não são adquiridas por interação direta. Por outro lado, ao agente fazer observações sobre mecânicos, emoções são vinculadas.

$J(\text{alice}, \text{john}, \text{mecânico}, \text{disponibilidade}, 0.9, 20/06/2015)$   
 $J(\text{alice}, \text{john}, \text{mecânico}, \text{qualidadeDoServiço}, 1, 20/06/2015)$   
 $J(\text{alice}, \text{john}, \text{mecânico}, \text{preço}, 0.8, 20/06/2015)$   
 $J(\text{alice}, \text{aaron}, \text{mecânico}, \text{disponibilidade}, 1, 15/02/2015)$   
 $J(\text{alice}, \text{aaron}, \text{mecânico}, \text{qualidadeDoServiço}, 0.5, 15/02/2015)$   
 $J(\text{alice}, \text{aaron}, \text{mecânico}, \text{preço}, -0.4, 15/02/2015)$   
 $J(\text{daniel}, \text{edward}, \text{mecânico}, \text{disponibilidade}, 1, 23/03/2014)$   
 $J(\text{daniel}, \text{edward}, \text{mecânico}, \text{qualidadeDoServiço}, -0.5, 23/03/2014)$   
 $J(\text{daniel}, \text{edward}, \text{mecânico}, \text{preço}, -0.4, 23/03/2014)$   
 $J(\text{bob}, \text{john}, \text{mecânico}, \text{disponibilidade}, 1,$   
 $\quad \{\text{concentrado} = 0.59\}, 30/06/2015)$   
 $J(\text{bob}, \text{aaron}, \text{mecânico}, \text{disponibilidade}, -1,$   
 $\quad \{\text{concentrado} = 0.6, \text{irritado} = 0.34\}, 30/06/2015)$   
 $J(\text{bob}, \text{edward}, \text{mecânico}, \text{disponibilidade}, 1,$   
 $\quad \{\text{concentrado} = 0.19\}, 30/06/2015)$

Após isso, estas crenças são enviadas como julgamentos para o contexto de crenças. Agora Bob tem um conjunto de julgamentos sobre outros mecânicos, que são enviadas ao contexto de confiança, onde o grau da confiança para cada agente é calculado. Emoções só influenciam na confiança direta, e, neste caso, esta é definida considerando apenas um julgamento, aquele sobre a disponibilidade do agente. Crenças de confiança são geradas e enviadas ao contexto de crenças. Neste momento, Bob pode selecionar entre suas novas crenças de confiança aquela que tem a melhor avaliação. A crença sobre John é a com melhor valor, e assim, ele é o mecânico escolhido.

$C(\text{john}, \text{mecânico}, \text{carro}(\text{funcionando}), 0.94, 30/06/2015)$

Após receber seu carro consertado, Bob gera um novo julgamento positivo.

$J(\text{bob}, \text{john}, \text{mecânico},$   
 $\text{disponibilidade}, 1, \{\text{feliz} = 0.41\}, 01/07/2015)$   
 $J(\text{bob}, \text{john}, \text{mecânico},$   
 $\text{qualidadeDoServiço}, 0.8, \{\text{feliz} = 0.41\}, 01/07/2015)$   
 $J(\text{bob}, \text{john}, \text{mecânico},$   
 $\text{preço}, 0.8, \{\text{feliz} = 0.41\}, 01/07/2015)$

Mais uma vez o carro de Bob está quebrado. Bob opta por não atualizar suas crenças de confiança sobre mecânicos. A crença de Steve é a melhor, porém, Steve não conseguirá atendê-lo. Então Bob escolhe a segunda melhor opção, John, para consertar seu carro. Chega um momento em que, antes de decidir qual mecânico escolher, Bob atualiza suas crenças de confiança, e percebe que John é mais confiável que Steve.

Os objetivos deste exemplo são: (i) descrever a dinâmica do agente; (ii) apresentar que falhas recentes podem comprometer a confiança; (iii) o agente não precisa atualizar suas crenças de confiança após um novo julgamento; e (iv) dependendo da quantidade e qualidade dos julgamentos, o agente pode dar peso maior para suas experiências ou para informações de terceiros por meio da função de reputação final (2.13).

## 5.2 TUTORIA

Este cenário simula um ambiente virtual de aprendizagem composto por dois tipos principais de entidades: repositórios e agentes tutores. Os repositórios disponibilizam cursos. Um mesmo curso pode ser fornecido por vários repositórios, com diferentes abordagens e custos. Cada agente tutor acompanha um aluno, e o direciona aos repositórios que satisfazem suas preferências. Agentes tutores podem trocar julgamentos sobre interações com repositórios.

É inicializado um agente no ambiente, *tutor1*, de acordo com o modelo proposto para acompanhar um aluno. Ele possui três desejos positivos, aqueles que ele gostaria que se tornassem reais, e um negativo, o qual ele rejeita.

$$(D^+(aprenderAutômatos), 0.8)$$

$$(D^+(aprenderMáquinasDeTuring), 0.6)$$

$$(D^+(exercíciosDeImplementação), 0.5)$$

$$(D^-(exercíciosDeDemonstração), 0.3)$$

As crenças iniciais do agente apresentam o seu conhecimento sobre os repositórios. Por exemplo, a crença

$$(B([interagirComRepositório1]aprenderAutômatos), 1)$$

representa que após a ação *interagirComRepositório1* ser executada, o agente acreditará que *aprenderAutômatos* é verdade. O grau 1 na crença representa a sua credibilidade. Por não preferir nenhum repositório, o grau de todas as crenças deste tipo recebem grau 1. Além disto, por não ter interagido com nenhuma entidade ainda, ele não possui julgamentos, nem crenças de confiança. As demais crenças do agente relacionadas à ações são:

$$(B([interagirComRepositório2]aprenderAutômatos), 1)$$

$$(B([interagirComRepositório3]aprenderAutômatos), 1)$$

$$(B([interagirComRepositório1]aprenderMáquinasDeTuring), 1)$$

$$(B([interagirComRepositório3]aprenderMáquinasDeTuring), 1)$$

As diretrizes sobre a dinâmica emocional são as mesmas apresentadas na seção 4.2. Para este agente em específico, enviar uma mensagem de solicitação, observar/agir de acordo com expectativas e receber mensagens com respostas geram impulsos positivos. Por outro lado, enviar uma mensagem fornecendo informação, observar/agir diferente do que o esperado e não receber mensagem no tempo estipulado geram impulsos negativos. A tabela 3 apresenta os valores dos estímulos emocionais a partir do tipo de evento que o provocou.

Tabela 3 – Valores para estímulos emocionais.

	Positivo	Negativo
Envio de mensagem	+10	-10
Observar/agir	+50	-50
Recebimento de mensagem	+10	-10

No estado inicial, o ponto de referência da dinâmica emocional está na posição  $(0, 0, 0)$ , implicando na localização no espaço PAD em  $(0, 0, 100)$ . Com isto, as emoções atuais são *concentrado* com intensidade 0.75, e *irritado* com intensidade 0.23.

No contexto de confiança estão representadas as seguintes informações em lógica descritiva:

- Descrição dos grupos: Existem dois grupos, um representando os tutores e coordenadores de tutoria que o agente conhece, e o outro com informações dos repositórios. Como apresentado na seção do modelo ReGreT (2.4.1.3), o cálculo da confiança considera os grupos dos agentes.
- Descrição dos domínios para realizar os desejos: cada desejo do agente pode incluir uma contextualização. Por exemplo, visando concretizar o desejo de *aprenderAutômatos*, o agente precisará fazer um curso fornecido por algum repositório. Com isto, informações sobre cursos e repositórios estão presentes na descrição do domínio.
- Viés: se o agente possuir mais informações externas do que internas, ele dará mais importância para a opinião de terceiros.
- Norma: se o agente receber julgamentos de um coordenador de tutoria, ele dará mais importância para estes julgamentos.

Algumas destas informações, como por exemplo a descrição do domínio, podem possuir a estrutura taxonômica compartilhada, onde o agente apenas adiciona as entidades de acordo com seus critérios. A figura 7 apresenta a estrutura básica destas informações.

Outras informações, representadas como indivíduos, que estão neste contexto são: o número de julgamentos que o modelo de confiança utiliza, quem são as entidades do mesmo grupo, entre outras.

No contexto de planejamento existe um conjunto de ações elementares que o agente pode executar. Por exemplo, ações visando observar uma entidade  $\varphi$ , ou enviar a mensagem  $\psi$  para uma entidade  $X$ .

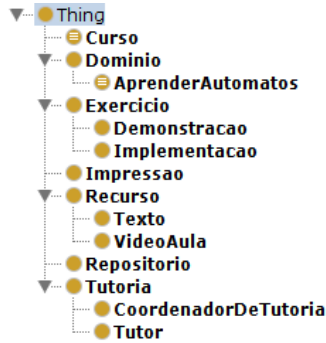
$$action(observar(\varphi), pre, observe(\varphi), c_\alpha)$$

$$action(enviarMensagem(X, \psi), pre, receive\ message(X, \psi), c_\alpha)$$

O agente não possui nenhuma intenção inicial, uma vez que elas são derivadas de crenças e desejos.



Figura 7 – Estrutura taxonômica presente no contexto de confiança.



### 5.2.1 Dinâmica do Agente

No início da execução, o agente escolhe o desejo com grau mais positivo para concretizá-lo.

$$(D^+(\text{aprenderAutômatos}), 0.8)$$

Ele é transferido de DC para PC com a regra de ponte 4.14. PC tem a tarefa de criar planos para concretizar este desejo. O conhecimento que está em BC e DC, necessário para o planejamento do desejo também é transferido pela regra de ponte 4.14. Ações, crenças e desejos são combinadas em PC na construção de planos. Por exemplo, a ação *interagirComRepositório1*, depende que o agente acredite na negação de *aprenderAutômatos*, deseje que *aprenderAutômatos* se torne verdade, e também que ele tenha uma crença de confiança sobre *repositório1*. O resultado desta ação é realizar seu desejo *aprenderAutômatos*.

```

action(
  interagirComRepositório1,
  belief(¬aprenderAutômatos, 1) ∧
  desire(aprenderAutômatos, 0.8) ∧
  confiança(repositório1, aprenderAutômatos, aprenderAutômatos, c, _),
  aprenderAutômatos, 1)

```

Como o agente *tutor1* não consegue satisfazer as precondições, ele precisa interagir com o ambiente na busca de informações. Isto é feito por meio da observação do *repositório1*, e do recebimento de mensagens sobre o desempenho desta entidade.

```

action(
  observar(repositório1),
  belief([interagirComRepositório1]aprenderAutômatos),
  observe(repositório1), 1)

action(
  enviarMensagem(X, repositório1),
  belief(tutoria(X)),
  receive message(X, repositório1), 1)

```

Estas três ações são combinadas em uma ação composta  $\alpha$ , com precondições *pre*, pós-condições *pos* e com confiança 1 neste plano, sendo que o objetivo deste plano é *aprenderAutômatos*. No conjunto de precondições, os desejos de *exercíciosDeImplementação* e *exercíciosDeDemonstração* devem ser considerados. Uma vez que estes são meios de realizar o objetivo de *aprenderAutômatos*. A forma como cada entidade apresenta conteúdo pode ser encontrada em seu contexto de confiança, onde existe a descrição dos repositórios.

```

best plan(
  aprenderAutômatos,  $\alpha$ , pre, pos, 1)

```

Este raciocínio aplicado para criação de planos para o *repositório1* é executado para os outros repositórios capazes de realizar o objetivo do agente de *aprenderAutômatos*, ou seja, *repositório2* e *repositório3*. Uma vez gerado um plano, a regra de ponte 4.15 é utilizada para criar uma intenção para este. Com isso, o agente *tutor1* tem um conjunto de intenções, visando realizar seu objetivo. Ele sempre escolhe a intenção de maior benefício para executar. A regra de ponte 4.16 é utilizada para enviar a CC a ação que será executada. Além disto, a intenção escolhida para ser executada é comunicada à EC pela regra de ponte 4.9, objetivando gerar estímulos emocionais.

Durante a execução dos planos, o agente *tutor1* recebe mensagens e faz observações. A regra de ponte 4.8 é utilizada para transferir estas per-

cepções de CC para EC. Uma vez em EC, observações fazem uso da regra de ponte 4.10 para virarem julgamentos, e se for o caso, vincular o estado emocional atual do agente.

### 5.2.1.1 Criação da Crença de Confiança

Supondo que o agente interagiu de modo a ter, em BC, todos os julgamentos necessários para o cálculo de confiança, a regra de ponte 4.12 transfere este conhecimento para TC. Assim, além dos julgamentos, TC faz uso das informações contidas no próprio contexto referentes a grupos, vieses e normas para o cálculo da confiança no modelo ReGreT.

De acordo com o modelo ReGreT, a reputação é dividida nas dimensões individual, social e ontológica. Estas três dimensões são combinadas gerando um valor que é vinculado a crença de confiança do agente. Assim, visando realizar o desejo de *aprenderAutômatos*, o agente cria crenças de confiança para cada um dos candidatos: *repositório1*, *repositório2* e *repositório3*.

Estas crenças de confiança são transferidas de TC para BC pela regra de ponte 4.13. Estando em BC, elas são transferidas para PC onde são vinculadas aos seus respectivos planos. E com isto, o planejamento escolhe interagir com uma das entidades visando *aprenderAutômatos*.

Os objetivos deste exemplo são: (i) descrever o raciocínio interno do agente, especificando as principais comunicações entre contextos; (ii) apresentar exemplos de propriedades que podem ser adicionadas ao contexto de confiança; e (iii) exemplificar a atribuição de valores para estímulos emocionais.

Neste capítulo foram apresentados dois cenários com o modelo de agente proposto. O primeiro, procurando um mecânico, descrevendo a execução em alto nível, e o segundo, tutoria, exemplificando a comunicação entre contextos. Este capítulo foi desenvolvido visando realizar o objetivo específico de avaliar o modelo proposto.



## 6 CONCLUSÃO

Neste trabalho foi definido um modelo de agentes *BDI-like*, em que confiança e emoções foram integradas ao raciocínio do agente. O modelo proposto seguiu a abordagem de sistemas multi-contexto, tendo como base o modelo BDI proposto por Casali, Godo e Sierra (2005). Ao modelo de Casali, Godo e Sierra foram adicionados dois contextos, um para confiança e outro para emoções. Regras de ponte foram criadas visando relacionar estes contextos com os demais componentes do agente. O modelo de confiança e reputação ReGreT (SABATER; SIERRA, 2001) foi adicionado. Para representar e manipular emoções foi utilizado o modelo WASABI (BECKER-ASANO; WACHSMUTH, 2010).

A escolha de SMC para representar o agente favoreceu a integração de confiança e emoções, uma vez que a correta definição de regras de ponte, permite a troca de conhecimento entre os componentes do agente. Esta abordagem foi utilizada objetivando integrar confiança em agentes BDI nos trabalhos de Pinyol e Sabater-Mir (2009) e Koster, Schorlemmer e Sabater-Mir (2012). Porém, nestes trabalhos emoções não foram consideradas.

A literatura de confiança e reputação apresenta um grande número de modelos, sendo a maior parte deles numéricos e alguns poucos cognitivos. A escolha por integrar um modelo numérico, ReGreT, foi motivada pelo trabalho de Bitencourt, Silveira e Marchi (2013), o qual relacionou emoções ao cálculo da confiança. Além disto, a integração de modelos cognitivos de confiança e reputação já foram analisadas em agentes BDI. O contexto de confiança é definido em lógica descritiva, permitindo a descrição dos domínios de forma não ambígua, podendo ser compartilhadas e integradas ao cálculo da confiança.

A interdisciplinaridade da área de emoções, e a existência de inúmeras definições para o termo, são fatores que influenciam a criação de vários modelos computacionais de emoções. Estes fatores motivaram a busca de um modelo de emoções que se adequasse a proposta deste trabalho. Embora foram apresentados dois modelos nesta pesquisa, outros modelos foram estudados. WASABI foi escolhido, principalmente, por representar informações de modo que a integração com confiança não precise de manipulações complexas.

Wooldridge (2009) aponta algumas diferenças existentes entre agentes e objetos. Entre elas, existe uma que afirma que um agente possui uma *thread* de controle. Objetos não necessariamente satisfazem esta condição. Visando reduzir o *overhead*, é assumido que cada contexto é uma *thread*. Por questões de escopo, uma ferramenta que implemente o modelo não foi criada. Desta

forma, apenas testes funcionais foram realizados.

## 6.1 TRABALHOS FUTUROS

A pesquisa focou na relação que emoções tem na definição de confiança. Desta forma, o impacto que emoções e confiança têm nos demais componentes do agente podem ser melhor explorados. A influência de normas no agente proposto também pode ser analisada.

A abordagem utilizada na definição da confiança foi baseada em crenças. A hipótese de marcadores somáticos de Damásio (DAMÁSIO, 1996) pode ser explorada para adicionar um aspecto não baseado em crenças. Para isso ser possível, é necessário adicionar a ideia de emoções secundárias no modelo. O modelo de emoções WASABI, pode ser utilizado para este propósito.

O modelo de confiança e reputação ReGret foi adaptado e adicionado ao agente proposto. Para adaptar o modelo ReGreT, uma função que integra confiança e emoções foi definida. Porém, por em cada julgamento utilizado pelo agente para definir a confiança estarem contidas, a avaliação racional e a influência emocional, a integração destes componentes com o modelo ReGreT pode ser melhor explorada. A influência de emoções em outros modelos numéricos, e em modelos cognitivos pode ser avaliada.

Por fim, esta pesquisa focou em teorias e arquiteturas de agentes, com isto, a implementação de uma ferramenta que possibilite a criação de agentes de acordo com o modelo proposto fica como trabalho futuro.

## REFERÊNCIAS

- BECKER-ASANO, C. **WASABI: Affect simulation for agents with believable interactivity**. [S.l.]: IOS Press, 2008.
- BECKER-ASANO, C. Wasabi for affect simulation in human-computer interaction. In: **Proc. International Workshop on Emotion Representations and Modelling for HCI Systems**. [S.l.: s.n.], 2014.
- BECKER-ASANO, C.; WACHSMUTH, I. Affective computing with primary and secondary emotions in a virtual human. **Autonomous Agents and Multi-Agent Systems**, Springer, v. 20, n. 1, p. 32–49, 2010.
- BIGA, A.; CASALI, A. G-jason: An extension of jason to engineer agents capable to reason under uncertainty. 2013.
- BITENCOURT, G. K.; SILVEIRA, R. A.; MARCHI, J. TrustE: An emotional trust model for agents. In: **EUMAS**. [S.l.: s.n.], 2013. p. 54–67.
- BONNEFON, J.-F. et al. A logical framework for trust-related emotions. **Electronic Communications of the EASST**, v. 22, 2010.
- BORDINI, R. H.; HÜBNER, J. F.; WOOLDRIDGE, M. **Programming multi-agent systems in AgentSpeak using Jason**. [S.l.]: John Wiley & Sons, 2007.
- BOSSE, T.; HÖHLE, D. Enhancing believability of virtual soccer players: Application of a bdi-model with emotions and trust. In: **Developing Concepts in Applied Intelligence**. [S.l.]: Springer, 2011. p. 119–128.
- BOSSE, T. et al. Leadsto: a language and environment for analysis of dynamics by simulation. In: **Multiagent System Technologies**. [S.l.]: Springer, 2005. p. 165–178.
- BRATMAN, M. **Intention, plans, and practical reason**. [S.l.]: Harvard University Press, 1987. ISBN 9780674458185.
- CASALI, A.; GODO, L.; SIERRA, C. Graded BDI models for agent architectures. In: **Computational Logic in Multi-Agent Systems**. [S.l.]: Springer, 2005. p. 126–143.
- CASALI, A.; GODO, L.; SIERRA, C. A graded BDI agent model to represent and reason about preferences. **Artificial Intelligence**, Elsevier, v. 175, n. 7, p. 1468–1478, 2011.

CASALI, A. et al. **On intentional and social agents with graded attitudes**. Tese (Doutorado), 2008.

CASTELFRANCHI, C.; FALCONE, R. Social trust: A cognitive approach. *Citeseer*, p. 55–90, 2001.

CASTELFRANCHI, C.; FALCONE, R. **Trust theory: A socio-cognitive and computational model**. [S.l.]: John Wiley & Sons, 2010.

CATADAU, I. P. **Milking the reputation cow: argumentation, reasoning and cognitive agents**. Tese (Doutorado) — Universitat Politècnica de Catalunya, 2010.

COHEN, P. R.; LEVESQUE, H. J. Intention is choice with commitment. **Artificial intelligence**, Elsevier, v. 42, n. 2, p. 213–261, 1990.

COMMERCE, B. E.; JØSANG, A.; ISMAIL, R. The beta reputation system. In: CITESEER. **In Proceedings of the 15th Bled Electronic Commerce Conference**. [S.l.], 2002.

CÔRTEZ, M. V. P. **UEBDI: Uma arquitetura para agentes emocionais BDI**. Dissertação (Mestrado) — Universidade Federal Fluminense, 2013.

DAMÁSIO, A. **O Erro de Descartes; trad. Dora Vicente e Georgina Segurado**. [S.l.]: São Paulo: Companhia das Letras, 1996.

EL-NASR, M. S.; YEN, J.; IOERGER, T. R. Flame - fuzzy logic adaptive model of emotions. **Autonomous Agents and Multi-agent systems**, Springer, v. 3, n. 3, p. 219–257, 2000.

FRANKLIN, S.; GRAESSER, A. Is it an agent, or just a program?: A taxonomy for autonomous agents. In: **Proceedings of the Workshop on Intelligent Agents III, Agent Theories, Architectures, and Languages**. London, UK, UK: Springer-Verlag, 1997. (ECAI '96), p. 21–35. ISBN 3-540-62507-0.

FULLAM, K. K. et al. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In: **ACM. Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems**. [S.l.], 2005. p. 512–518.

GAMBETTA, D. et al. Can we trust trust. **Trust: Making and breaking cooperative relations**, v. 2000, p. 213–237, 2000.



- GEBHARD, P. Alma: a layered model of affect. In: **ACM. Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems**. [S.l.], 2005. p. 29–36.
- GEORGEFF, M. P.; LANSKY, A. L. Procedural knowledge. **Proceedings of the IEEE**, IEEE, v. 74, n. 10, p. 1383–1398, 1986.
- GIUNCHIGLIA, F.; SERAFINI, L. Multilanguage hierarchical logics, or: how we can do without modal logics. **Artificial intelligence**, Elsevier, v. 65, n. 1, p. 29–70, 1994.
- HAY, L. S. Axiomatization of the infinite-valued predicate calculus. **The Journal of Symbolic Logic**, Cambridge Univ Press, v. 28, n. 01, p. 77–86, 1963.
- HERNANDEZ, D. et al. BDIE: a BDI like architecture with emotional capabilities. In: **American Association for Artificial Intelligence Spring Symposium**. [S.l.: s.n.], 2004.
- HERZIG, A. et al. A logic of trust and reputation. **Logic Journal of IGPL**, Oxford Univ Press, v. 18, n. 1, p. 214–244, 2010.
- HOELZ, B.; RALHA, C. Towards a cognitive meta-model for adaptive trust and reputation in open multi-agent systems. **Autonomous Agents and Multi-Agent Systems**, Springer US, p. 1–32, 2014. ISSN 1387-2532.
- HÜBNER, J. F. et al. From cognitive trust theories to computational trust. In: **Proceedings of the 12th International Workshop on Trust in Agent Societies, Budapest, Hungary**. [S.l.: s.n.], 2009. v. 10, n. 05, p. 2009–11.
- HUYNH, T. D.; JENNINGS, N. R.; SHADBOLT, N. R. An integrated trust and reputation model for open multi-agent systems. **Autonomous Agents and Multi-Agent Systems**, Springer, v. 13, n. 2, p. 119–154, 2006.
- JENNINGS, N. R.; SYCARA, K.; WOOLDRIDGE, M. A roadmap of agent research and development. **Autonomous agents and multi-agent systems**, Kluwer Academic Publishers, v. 1, n. 1, p. 7–38, 1998.
- JIANG, H.; VIDAL, J. M.; HUHNS, M. N. EBDI: an architecture for emotional agents. In: **ACM. Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems**. [S.l.], 2007. p. 11.
- JØSANG, A.; ISMAIL, R.; BOYD, C. A survey of trust and reputation systems for online service provision. **Decision support systems**, Elsevier, v. 43, n. 2, p. 618–644, 2007.

KERR, R.; COHEN, R. Treet: The trust and reputation experimentation and evaluation testbed. **Electronic Commerce Research**, Springer, v. 10, n. 3-4, p. 271–290, 2010.

KOLLINGBAUM, M. J. **Norm-governed practical reasoning agents**. Tese (Doutorado) — University of Aberdeen, 2005.

KOSTER, A.; SCHORLEMMER, M.; SABATER-MIR, J. Opening the black box of trust: reasoning about trust models in a BDI agent. **Journal of Logic and Computation**, 2012.

LEJMI-RIAH, H.; KEBAIR, F.; SAID, L. B. Agent decision-making under uncertainty: Towards a new E-BDI agent architecture based on immediate and expected emotions. **International Journal of Computer Theory & Engineering**, v. 6, n. 3, 2014.

ŁUKASIEWICZ, J. O logice trójwartościowej. 1920.

MARSELLA, S.; GRATCH, J.; PETTA, P. Computational models of emotion. **A Blueprint for Affective Computing-A sourcebook and manual**, p. 21–46, 2010.

MARSELLA, S. C.; GRATCH, J. Ema: A process model of appraisal dynamics. **Cognitive Systems Research**, Elsevier, v. 10, n. 1, p. 70–90, 2009.

MEHRABIAN, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. **Current Psychology**, Springer, v. 14, n. 4, p. 261–292, 1996.

MEYER, J.-J. C. et al. Logics for intelligent agents and multi-agent systems. Elsevier, 2014.

MISZTAL, B. **Trust in modern societies: The search for the bases of social order**. [S.l.]: John Wiley & Sons, 2013.

NORIEGA, P.; SIERRA, C. Towards layered dialogical agents. In: MÜLLER, J.; WOOLDRIDGE, M.; JENNINGS, N. (Ed.). **Intelligent Agents III Agent Theories, Architectures, and Languages**. [S.l.]: Springer Berlin Heidelberg, 1997, (Lecture Notes in Computer Science, v. 1193). p. 173–188. ISBN 978-3-540-62507-0.

ORTONY, A.; CLORE, G.; COLLINS, A. **The Cognitive Structure of Emotions**. [S.l.]: Cambridge University Press, 1990. ISBN 9780521386647.

PARSONS, S.; SIERRA, C.; JENNINGS, N. Agents that reason and negotiate by arguing. **Journal of Logic and computation**, Oxford Univ Press, v. 8, n. 3, p. 261–292, 1998.

PEREIRA, D. et al. Towards an architecture for emotional bdi agents. In: **In EPIA '05: Proceedings of 12th Portuguese Conference on Artificial Intelligence**. [S.l.]: Springer, 2005. p. 40–47.

PICARD, R. W.; PICARD, R. **Affective computing**. [S.l.]: MIT press Cambridge, 1997.

PINYOL, I.; SABATER-MIR, J. Pragmatic-strategic reputation-based decisions in BDI agents. In: INTERNATIONAL FOUNDATION FOR AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS. **Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2**. [S.l.], 2009. p. 1001–1008.

PINYOL, I.; SABATER-MIR, J. Computational trust and reputation models for open multi-agent systems: a review. **Artificial Intelligence Review**, Springer, v. 40, n. 1, p. 1–25, 2013.

PUICĂ, M.-A.; FLOREA, A.-M. Emotional belief-desire-intention agent model: Previous work and proposed architecture. **International Journal of Advanced Research in Artificial Intelligence**, Citeseer, v. 2, n. 2, p. 1–8, 2013.

RAO, A. S.; GEORGEFF, M. P. Modeling rational agents within a BDI-architecture. **KR**, v. 91, p. 473–484, 1991.

RAO, A. S.; GEORGEFF, M. P. et al. BDI agents: From theory to practice. In: **ICMAS**. [S.l.: s.n.], 1995. v. 95, p. 312–319.

RODRÍGUEZ, L.-F.; RAMOS, F. Development of computational models of emotions for autonomous agents: a review. **Cognitive Computation**, Springer, v. 6, n. 3, p. 351–375, 2014.

ROUSSEAU, D. M. et al. Not so different after all: A cross-discipline view of trust. **Academy of management review**, Academy of Management, v. 23, n. 3, p. 393–404, 1998.

RUSSELL, J. A.; MEHRABIAN, A. Evidence for a three-factor theory of emotions. **Journal of research in Personality**, Elsevier, v. 11, n. 3, p. 273–294, 1977.

RUSSELL, S.; NORVIG, P.; SOUZA, V. D. de. **Inteligência artificial: tradução da segunda edição**. [S.l.]: Elsevier: Campus, 2004.

SABATER, J.; PAOLUCCI, M.; CONTE, R. Repage: Reputation and image among limited autonomous partners. **Journal of artificial societies and social simulation**, v. 9, n. 2, 2006.

SABATER, J.; SIERRA, C. ReGreT: reputation in gregarious societies. In: ACM. **Proceedings of the fifth international conference on Autonomous agents**. [S.l.], 2001. p. 194–195.

SABATER, J.; SIERRA, C. Review on computational trust and reputation models. **Artificial intelligence review**, Springer, v. 24, n. 1, p. 33–60, 2005.

SABATER, J. et al. Using multi-context systems to engineer executable agents. In: **Intelligent Agents VI. Agent Theories, Architectures, and Languages**. [S.l.]: Springer, 2000. p. 260–276.

SABATER-MIR, J. **Trust and Reputation for Agent Societies**. Tese (Doutorado), 2002.

SABATER-MIR, J.; PAOLUCCI, M. On representation and aggregation of social evaluations in computational trust and reputation models. **International Journal of Approximate Reasoning**, Elsevier, v. 46, n. 3, p. 458–483, 2007.

SABATER-MIR, J.; SIERRA, C. **Agreement Technologies**. [S.l.]: Springer Netherlands, 2012. (Law, Governance and Technology Series). ISBN 9789400755833.

SCHERER, K. R. Appraisal theory. **Handbook of cognition and emotion**, p. 637–663, 1999.

SCHWARZ, N.; CLORE, G. L. How do i feel about it? the informative function of affective states. **Affect, cognition, and social behavior**, p. 44–62, 1988.

SIGNORETTI, A. **Agentes Inteligentes com Foco de Atenção Afetivo em Simulações Baseadas em Agentes**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2012.

SMITH, N. K. et al. Being bad isn't always good: Affective context moderates the attention bias toward negative information. **Journal of Personality and Social Psychology**, American Psychological Association, v. 90, n. 2, p. 210, 2006.

STEUNEBRINK, B. R.; DASTANI, M.; MEYER, J.-J. C. Towards a quantitative model of emotions for intelligent agents. In: **Proceedings of the**

**2nd Workshop on Emotion and Computing-Current Research and Future Impact, Osnabrück, Germany.** [S.l.: s.n.], 2007.

STEUNEBRINK, B. R.; DASTANI, M.; MEYER, J.-J. C. The OCC model revisited. In: **Proc. of the 4th Workshop on Emotion and Computing.** [S.l.: s.n.], 2009.

TAKEDA, H.; VEERKAMP, P.; YOSHIKAWA, H. Modeling design process. **AI magazine**, v. 11, n. 4, p. 37, 1990.

WOBCKE, W. Reasoning about BDI agents from a programming languages perspective. In: **AAAI Spring Symposium: Intentions in Intelligent Systems.** [S.l.]: AAAI, 2007. p. 48–.

WOOLDRIDGE, M. **Reasoning about Rational Agents.** [S.l.]: MIT Press, 2000. ISBN 9780262265027.

WOOLDRIDGE, M. Intelligent agents: The key concepts. In: **Multi-Agent Systems and Applications II.** [S.l.]: Springer, 2002. p. 3–43.

WOOLDRIDGE, M. **An introduction to multiagent systems.** [S.l.]: John Wiley & Sons, 2009.

WOOLDRIDGE, M.; JENNINGS, N. R. Intelligent agents: Theory and practice. **The knowledge engineering review**, Cambridge Univ Press, v. 10, n. 02, p. 115–152, 1995.