



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



CULTURA
ACADÊMICA
Editora

Motivação e emoções em criaturas naturais e artificiais

Ricardo Ribeiro Gudwin

Como citar: GUDWIN, R. R. Motivação e emoções em criaturas naturais e artificiais. *In:* ALVES, M. A. (org.). **Cognição, emoções e ação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2019. p. 151-176.
DOI: <https://doi.org/10.36311/2019.978-85-7249-019-1.p151-176>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

MOTIVAÇÃO E EMOÇÕES EM CRIATURAS NATURAIS E ARTIFICIAIS

Ricardo R. Gudwin

Universidade Estadual de Campinas - Unicamp
gudwin@g.unicamp.br

INTRODUÇÃO

Em computação, o estudo de *Criaturas Artificiais*, tal como realizado, por Balkenius (1995), por exemplo, diz respeito ao desenvolvimento de personagens sintéticos, seres artificiais que imitam, de alguma forma, seres vivos, com os mais diversos propósitos. Dentre estes propósitos, encontramos o estudo dos princípios basilares da vida, estudados na área de *Vida Artificial (Artificial Life)* por pesquisadores como Steels (1993), a criação de sistemas de apoio ao desenvolvimento humano (sistemas tutores, sistemas de cuidados com idosos ou pessoas adoentadas, sistemas de auxílio na divulgação de informações etc.) ou até mesmo o desenvolvimento de sistemas dedicados ao puro e simples entretenimento (como no caso dos jogos digitais).

O estudo de criaturas artificiais pode ser visto como uma sub-área da pesquisa em agentes inteligentes, realizado por Wooldridge e Jennings (1995) e Franklin (1995), dentre outros. Nesse contexto, uma criatura artificial é definida como um agente inteligente que possui algum tipo de “corpo” (real ou virtual), que lhe dá identidade¹ e serve como interface para a criatura artificial interagir com seu ambiente. Dessa forma, o estudo sobre criaturas artificiais abrange tanto a robótica autônoma² quanto a vida artificial, incluindo-se aí os jogos de computador, onde personagens

¹ Essa questão do corpo pode ser bastante problemática, como mostram Carruthers (2008), Oka *et.al.* (2001) e Wilson (2002).

² Entende-se por “robótica autônoma” o ramo da robótica em que as ações realizadas pelos robôs não são comportamentos fixos pré-programados, mas o robô deve sernsorear o ambiente e, em função de seus objetivos, decidir por seus próprios meios uma sequência de ações que deve ser executada sobre o ambiente procurando atingir esses objetivos, sendo que essas ações podem ser fruto de experimentação sobre o ambiente e aprendizado auto-organizado.

<https://doi.org/10.36311/2019.978-85-7249-019-1.p151-176>

sintéticos podem ser utilizados tanto como colaboradores quanto como oponentes inteligentes. A inspiração para a construção de criaturas artificiais tanto pode vir de estudos do comportamento animal (caso em que essas criaturas artificiais são chamadas por Dean (1998) de *animats*, como do comportamento humano, ambos estudados pelas ciências cognitivas.

O uso de modelos cognitivos³ como fonte de inspiração para o desenvolvimento de sistemas de controle para criaturas artificiais deu origem à área de Mentes Artificiais (Artificial Minds), estudadas por Franklin (1995), por exemplo, construídas com base em Arquiteturas Cognitivas, conforme Langley, Laird e Rogers (2009). Essas mentes artificiais são projetadas seguindo arquiteturas computacionais construídas lastreadas em modelos cognitivos. Estes têm como base abordagens advindas da filosofia da mente, da psicologia, da neurociência, da inteligência artificial, da linguística, da etologia, da antropologia e demais ciências cognitivas. Muitas vezes, os pesquisadores da área adotam uma terminologia oriunda dessas áreas e passam a empregá-la de maneira ingênua aos modelos computacionais que desenvolvem, sem maiores cuidados no uso de termos que foram originalmente cunhados para designar funcionalidades e elementos da mente humana (ou animal), causando grande desconforto em estudiosos em ciências cognitivas. Outras vezes, desenvolvem uma transposição cuidadosa dos modelos cognitivos, originalmente desenvolvidos para explicar características da mente humana, efetuando abstrações cautelosas de tal forma que os termos empregados possam dizer respeito tanto a mentes humanas como a mentes artificiais.

Um dos termos que causa grande desconforto, quando aplicado de maneira ingênua a mentes artificiais, é exatamente o termo “emoção”. Para muitos pesquisadores, emoções são uma característica muito peculiar aos seres humanos e outros animais. Ela jamais poderia ser utilizada para descrever componentes de mentes sintéticas, criadas pelo ser humano. Nesta linha de raciocínio, as emoções e os sentimentos seriam uma característica intrínseca de nossa humanidade, que nos permitiriam experienciar o mundo de maneira consciente, sem jamais poder ser atribuídas a seres artificiais.

³ Entende-se aqui os modelos conceituais, matemáticos-computacionais ou processuais de funcionalidades cognitivas, que permitam uma instanciação em sistemas artificiais, descartando-se outros tipos de modelos, tais como modelos animais, por exemplo.

Mas, seria este mesmo o caso? Afinal de contas, o que são de fato emoções? Todos nós sabemos o que é sentir medo, alegria, felicidade, a emoção de um primeiro amor, de conseguir o primeiro emprego, de se sentir rejeitado ou sofrer a reprimenda por algo feito errado. Mas o que é de fato isso? Um estado? Um processo? Uma função?

Apesar de todo ser humano ter uma ideia intuitiva do que é uma emoção, as tentativas de se explicar formalmente a sua natureza são muitas e diversas, de modo que carecemos de uma definição clara e consensual do que sejam emoções⁴. Se procurarmos na literatura especializada, veremos que o tópico emoções é estudado em diversas áreas do conhecimento. Há diversos modelos, normalmente incompatíveis entre si, tentando explicar o fenômeno, conforme Cabanac (2002) e Izard (2010). Ademais, o termo emoção está intimamente ligado a diversos outros termos correlatos, tais como afetos, sentimentos, desejos, impulsos, vontades, *drives*, motivações, necessidades e assim por diante.

Se, de uma maneira ingênua, procurarmos o verbete “emoções” na Wikipédia, por exemplo, encontramos a definição de emoção como sendo qualquer experiência consciente caracterizada por intensa atividade mental e alto grau de prazer ou desprazer, com a ressalva de que o discurso científico está evitado de outros significados para o termo, e que não existe consenso sobre essa definição. Nesse contexto, o conceito é frequentemente associado a termos correlatos, tais como “humor”, “temperamento”, “personalidade”, “disposição” ou “motivação”.

Uma busca pela origem etimológica do termo mostra que emoção vem de *e*, variante de *ex*, mais *movere*, que significa por para fora, ou seja, a etimologia está associada à expressão das emoções, aos caracteres fisiológicos usualmente associados à expressão das emoções, mas não nos dá uma pista adequada do que é o fenômeno internamente, do ponto de vista cognitivo.

Apesar de um grande número de modelos aparentemente divergentes entre si, como mostra Johnson (2009), o uso de emoções em sistemas inteligentes e particularmente aplicado ao desenvolvimento de criaturas artificiais possui diversas contribuições interessantes e importantes

⁴ Izard (2010) relata a compilação de nada menos do que 92 diferentes definições para o termo “emoção”, coletadas a partir de livros e artigos em periódicos.

(BATES, 1994; BLUMBERG, 1996; CAÑAMERO, 2005; EL-NASR, 1998; KODA, 1996; MALFAZ, 2004; MEYER, 2006; ORTONY, 2003; REILLY, 1996; SARMENTO, 2004; TOMLINSON; BLUMBERG, 2002; VELASQUEZ, 1998; VENTURA, 2000; ZIEMKE, 2009).

Neste capítulo, fazemos uma revisão de diferentes modelos computacionais de emoções, apontando suas similaridades e diferenças, bem como sua associação direta com a questão das motivações. Apresentamos a diversidade de abordagens e sinalizamos a direção que, em nosso ponto de vista, mostra-se a mais adequada ao uso não ingênuo do termo, segundo as tendências da ciência cognitiva. Ao final, desenvolvemos uma proposta que visa unificar essas diferentes abordagens, buscando contemplar todas as posições apontadas no presente estudo.

1 PRECURSORES DO ESTUDO DE EMOÇÕES EM SISTEMAS INTELIGENTES

Os precursores mais antigos no estudo das emoções vêm da filosofia, onde surge a distinção entre ação e paixão, o estado de se estar ativo ou passivo, conforme mostram James (1997) e Dixon (2003). Nesse contexto, denominam-se paixões como os efeitos causados sobre a alma humana por elementos externos ao ser. Diversos filósofos, tais como Aristóteles, São Tomás de Aquino, Descartes, Hobbes e Spinoza, para citar alguns, debruçaram-se sobre o tema. De acordo com Rorty (1982), o conceito de paixão foi gradativamente sendo refinado. Os conceitos foram sendo refinados, de forma a se chegar ao conceito moderno de emoção empregado em psicologia. A partir da ideia de paixão como uma mera passividade a algo externo que invade nossa mente, como na perspectiva cartesiana, por exemplo, passa-se ao conceito de sentimento, de prazer e desprazer como reações avaliativas a possíveis diferentes cursos de ações. De uma mera passividade, as emoções transformam-se em uma ação da mente orientadora do comportamento inteligente, passando a ter um papel importante no processo da racionalidade, da qual se colocava em oposição originalmente⁵.

Dixon (2012) defende a ideia de que o conceito moderno de “emoção”, estudado na psicologia, surge a partir do século XIX como uma abstração para diversos conceitos relacionados, tais como apetites,

⁵ Veja o conceito de Inteligência Emocional, defendido por Goleman (1995), onde as emoções são apontadas como elemento fundamental no processo da racionalidade.

paixões, afetos e sentimentos, previamente investigados no âmbito da filosofia. Segundo Dixon, esse conceito moderno de emoção foi originalmente criado pelo pensador escocês Thomas Brown (1778-1820), sendo depois utilizado por Darwin e William James.

Cornelius (2000) apresenta as 4 principais perspectivas teóricas no estudo moderno das emoções:

A primeira perspectiva é oriunda de Darwin (1965) e apresenta as emoções como o resultado de um processo evolutivo. Nele, a seleção natural foi gradativamente dando forma a um conjunto de expressões comportamentais com funções bastante características e importantes no processo de garantir a sobrevivência de indivíduos de diferentes espécies animais, interagindo de diferentes maneiras em um ecossistema.

A segunda perspectiva é originada de James (1884), para o qual as emoções são a percepção de mudanças em nosso corpo, causando sentimentos, perspectiva também compartilhada por Damásio (1994, 1999, 2003). Por um lado, Darwin se ocupa mais com a questão da expressão das emoções, documentadas de maneira primorosa em seu trabalho sobre a expressão das emoções no homem e nos animais. Por outro lado, James ocupa-se em detalhar mais a experiência emocional da modificação do corpo, precedendo uma predisposição comportamental, que é vista como uma tendência a ações pré-determinadas, como consequência do processo evolutivo (em consonância com Darwin, nesse caso). Dessa forma, para James (e também para Damásio), a emoção estaria ligada às mudanças corporais, em um primeiro momento, seguido posteriormente do sentimento, visto como o resultado cognitivo da emoção, e causando uma determinação comportamental.

Ainda segundo Cornelius (2000), a terceira perspectiva é denominada cognitiva. É iniciada no trabalho de Arnold (1960) e talvez seja atualmente a vertente mais influente nos estudos sobre emoções em psicologia. A pressuposição central dessa perspectiva é a de que pensamentos e emoções são processos inseparáveis. Mais especificamente, as emoções são vistas como avaliadores (*appraisals*), ou seja, o processo por meio do qual eventos do ambiente são julgados bons ou ruins para um indivíduo. Arnold acusa James de excesso de simplicidade em dizer que as emoções são somente a percepção de mudanças corporais. De acordo com Arnold,

James falhou em perceber esse aspecto avaliativo (ou *appraisivo*) das emoções, que as distinguiria de uma mera percepção. Da mesma maneira que James não podia conceber as emoções sem as mudanças corporais, Arnold não podia conceber emoções sem esse caráter avaliativo.

Finalmente, a quarta perspectiva sobre as emoções, mais recente e controversa, é a chamada perspectiva social-construtivista. Contrariamente às perspectivas mais biológicas, que defendem as emoções como uma adaptação oriunda de um processo evolutivo, a perspectiva social-construtivista acredita que as emoções são produtos culturais da aprendizagem de regras sociais. Segundo seus defensores, para se entender todo o espectro de complexidades envolvidas nas emoções, é necessário ir além do aspecto evolutivo e adentrar nos aspectos socioculturais envolvendo as interações humanas e suas regras sociais.

Estas quatro perspectivas implicitamente sugerem diferentes modelos do fenômeno emocional, nos apontando diferentes aspectos importantes na modelagem das emoções. Essa modelagem das emoções não fica somente na questão da importância das mudanças no corpo (expressão das emoções), ou na questão do *appraisal*⁶ impactando a tomada de decisões e comportamento. Alguns autores começaram a detectar outras dimensões que pareciam ser importantes na compreensão do fenômeno emocional. Schlosberg (1954), por exemplo, aponta a importância do que ele chama de “nível de ativação”, ou intensidade de uma emoção, que autores posteriores iriam identificar como o caráter do *arousal*, que diferencia emoções “fortes” de emoções “fracas”. A partir dessas duas dimensões de *appraisal* e *arousal*, foram desenvolvidos modelos de emoções como o PAT (Pleasure-Arousal Theory) de Reisenzein (1994) ou o Modelo Circumplexo de afeto, exposto em Russell (1980) e em Posner, Russell e Peterson (2005), além de outros, como o de Barrett (1998).

Outros autores passaram a incorporar ainda outras dimensões além das citadas acima. Dentre elas, o eixo dominância-submissividade, conforme Russell (1977) e Mehrabian (1996), ou até mesmo a duração temporal e uma polêmica dimensão de “qualidade”, conforme Cabanac

⁶ Apesar de podermos utilizar aqui a palavra em português “avaliador”, entendemos que o termo original em inglês “appraisal” talvez seja mais significativo que o termo “avaliador” em português. Desta forma, manteremos o termo original em inglês ao invés de adotar uma versão em português que talvez não expresse completamente a ideia original.

(2002), possivelmente presumida diante da constatação de que emoções como o medo e a fome possuiriam diferentes qualidades. A questão mal resolvida do que seria essa “qualidade” começa a se solucionar, quando se passa a diferenciar uma única dimensão hedônica de prazer-desprazer, em múltiplas dimensões avaliativas. O termo *appraisal* passa a ser entendido como um processo que detecta e avalia a significância de diversos fatores para o bem estar de um indivíduo. Esse conceito de “bem estar do indivíduo” passa a incorporar a satisfação ou obstrução de potencialmente múltiplas necessidades, valores, metas e crenças do indivíduo. Em outras palavras, quer dizer, qualquer coisa com que um indivíduo pode se importar, conforme Moors (2013).

Essa última evolução do modelo de emoções aponta uma conexão óbvia com o modelo de comportamento motivacional desenvolvido por Hull (1952), que passa a ter uma grande importância no desenvolvimento de modelos de emoções em criaturas artificiais. Segundo Hull, o comportamento motivado pode ser explicado por meio de um conjunto de *necessidades*, que direcionam uma ação correspondente no ambiente. Está implícito no modelo que esta ação deve causar uma redução nessa necessidade ou, de outra forma, que essa necessidade deve ser satisfeita a partir dos resultados causados por esta ação. Hull desenvolve, então, o conceito de *drive*, como uma medida que avalia o quanto uma determinada necessidade foi satisfeita. Um *drive* alto indica que a necessidade possui um baixo nível de satisfação, causando maior impacto nas decisões comportamentais. Um *drive* baixo indica que a necessidade foi satisfatoriamente saciada. Segundo Hull, um organismo vivo pode possuir todo um conjunto de diferentes necessidades, cada uma delas com um nível diferente de intensidade (ou de satisfação), a cada instante de tempo. O papel do sistema motivacional é avaliar o estado de satisfação dessas necessidades, mensurado pelo nível de intensidade dos *drives* correspondentes, e escolher a melhor ação no sentido de satisfazer às necessidades que possuem *drives* mais intensos.

Os *drives* podem ser fisiológicos ou sociais. Necessidades fisiológicas dão origem aos assim chamados *drives* primários de baixo nível. Exemplos deste tipo apontados na literatura são fome, sede, sono e segurança (preservação de danos físicos). Necessidades sociais dão origem aos *drives* primários de alto nível. Exemplos desse tipo incluem a amizade, o

desejo por pertencimento social, por reciprocidade, a curiosidade, a autonomia e a manutenção da honra.

Alguns trabalhos na literatura relatam uma hierarquia de necessidades, como o trabalho de Maslow (1943). Nessa hierarquia, ao invés de somente dois níveis de prioridade (onde *drives* de baixo nível são mais prioritários em relação a *drives* de alto nível), diferentes níveis hierárquico podem ser estabelecidos, criando uma hierarquia de prioridades entre necessidades. A literatura ainda aponta a existência de *drives* secundários que, ao contrário de *drives* primários, determinados somente a partir de informações sensoriais na avaliação da satisfação de necessidades, avaliam essa satisfação de necessidades a partir da composição de informações sensoriais e de outros *drives*. Dessa forma, *drives* secundários são aqueles que são originados a partir de outros *drives*.

Fica clara, assim, a conexão entre teorias sobre emoções com teorias sobre motivações, conforme Buck (1988). Na verdade, em alguns casos, é até difícil distinguir onde emoções e motivações se diferenciam, pois, de fato, os temas são bastante interligados. Necessidades, *drives*, motivações e emoções parecem ser componentes de um modelo complexo de sistemas motivacionais, que podem ser uma parte importante na determinação do comportamento humano e animal.

Por fim, vale a pena apontar a importância das emoções no comportamento racional, apontada mais recentemente nos trabalhos de Damásio (1994, 1999, 2003) e Goleman (1995). Os autores defendem que o comportamento racional depende diretamente do caráter apraisivo das emoções. Sem as emoções como critério, a racionalidade seria impossível, segundo defendem esses autores.

Nas próximas seções, apresentamos alguns dos modelos desenvolvidos utilizando tanto teorias sobre emoções como teorias sobre motivações em criaturas artificiais.

2 MODELO OCC: ORTONY, CLORE, COLLINS

Um dos modelos clássicos das emoções, originado a partir da psicologia cognitiva, mas com grande impacto no desenvolvimento de sistemas de cognição artificial é o modelo desenvolvido por Ortoni, Clore

e Collins (1988), sobre a estrutura cognitiva das emoções, chamado aqui de modelo OCC. Trata-se de um dos típicos modelos originados a partir da dicotomia *appraisal / arousal*, chamadas pelos autores de *valência e ativação*. Segundo o modelo OCC, emoções são estados mentais internos que podem ter diferentes intensidades e que estão predominantemente relacionados a afetos. Estes, por sua vez, são reações avaliativas a situações determinadas como boas ou ruins, podendo ser reações ponderadas a:

- **Eventos** e suas consequências (coisas que acontecem e que são percebidas pelo indivíduo);
- **Agentes** e suas ações (outras pessoas ou seres animados que atuam no mundo - incluindo-se uma estimativa dos estados mentais desses outros agentes, dando lugar a emoções tais como orgulho, admiração, vergonha, censura);
- **Objetos** e suas propriedades (emoções surgidas das percepções sobre o objeto).

O modelo OCC desenvolve-se na especificação de 22 emoções implementadas em termos de variáveis locais e globais. Um dos diferenciais de OCC é justamente o fato de empregar ferramentas da inteligência artificial na modelagem da teoria. Ele pode ser representado na forma de regras, que consideram o potencial detonador da emoção, o limiar para a detonação da emoção e a intensidade para a emoção. O potencial da emoção pode ser visto como um desejo (de uma consequência de um evento), um merecimento (de uma ação em um agente) ou um apelo (de um objeto).

A distinção entre reações a eventos, agentes e objetos, característica do modelo OCC, nos leva portanto a distinguir três tipos de emoções:

- Ficar contente ou descontente (com a ocorrência de um evento - reações a eventos)
- Reações de aprovação ou desaprovação (a ações realizadas por outros agentes - reações a agentes)
- Gostar ou desgostar (de um determinado objeto ou estado - reações a objetos)

O modelo OCC é um meta-modelo⁷ bastante citado em diversos trabalhos envolvendo criaturas artificiais e a síntese de emoções artificiais (BATES, 1994; EL-NASR, 1998; EL-NASR; YEN; IOERGER, 2000; KODA, 1996; REILLY, 1996; VELÁSQUEZ, 1998; WALLACH; ALLEN, 2008).

3 DAMÁSIO E O MARCADOR SOMÁTICO

O trabalho de Damásio (1994, 1999, 2003) é outro bastante citado em diversos artigos envolvendo emoções em sistemas inteligentes, como, por exemplo, Velásquez (1998, 1999), Ventura (2000), Tomlinson e Blumberg (2002), Malfaz e Salichs (2004), Sarmento (2004), Ziemke e Lowe (2009). Em “O Erro de Descartes”, Damásio (1994) propõe a racionalidade com emoção e a necessidade de um corpo no processo. O autor propõe uma distinção entre emoção e sentimento, propondo que uma *emoção* seria uma modificação no corpo em função de estímulos externos (seguindo basicamente as ideias de William James) e os sentimentos seriam uma contrapartida cognitiva das emoções, surgindo em função da tomada de consciência delas.

Damásio classifica as emoções em Primárias e Secundárias. As Primárias são aquelas que experienciamos desde a infância, para as quais se supõe haver um mecanismo pré-organizado, enquanto as Secundárias são próprias de um indivíduo adulto. Dentre as primárias estariam o medo, alegria, tristeza, raiva. Já dentre as emoções secundárias encontramos o ciúme, culpa, orgulho. Emoções primárias são normalmente disparadas a partir de estímulos externos ou internos, a partir dos diversos órgãos sensoriais, ao passo que emoções secundárias seriam disparadas puramente a partir de eventos cognitivos.

Damásio elabora, então, sua hipótese do marcador somático, onde as emoções funcionam como marcas de certos aspectos de situações ou o resultado de possíveis ações, que podem ser utilizadas como elementos diferenciais no processo de tomada de decisão. Quando experienciamos uma emoção, o resultado imediato é uma alteração no corpo, tal como a aceleração do batimento cardíaco ou o aumento da sudorese. A essa emoção

⁷ Um meta-modelo é um modelo de um modelo, ou seja, um tipo de modelo abstrato que demanda ser instanciado em um modelo concreto de forma a poder ser utilizado.

estaria associada um sentimento, uma avaliação cognitiva sobre a conveniência ou não da situação para o indivíduo. Essa alteração corporal marca a situação que a causou. Com isso, outras situações que causem situações semelhantes serão classificadas, de modo análogo, como boas ou ruins. Isso irá impactar a tomada de decisões subsequentes em função das novas alterações corporais evocadas. Além disso, propõe que os sistemas cerebrais que participam conjuntamente da emoção e da tomada de decisões estão envolvidos na gestão da cognição e do comportamento social.

As emoções, neste contexto, funcionariam tanto como um critério inicial na tomada de decisões, onde um raciocínio elaborado ainda não foi desenvolvido (uma espécie de razão inconsciente, instintiva, quase que reativa, mas claramente motivacional), como na tomada de decisões que envolve uma deliberação mais aprofundada. No primeiro caso, dependendo da intensidade da alteração corporal, a decisão pode ser visceral, definitiva. No segundo caso, normalmente diferentes possíveis cenários são avaliados. Mesmo assim, o critério de julgamento se define a partir de algum *appraisal* advindo de uma memória emocional.

4 SIMON E SLOMAN: EMOÇÕES COMO FILTROS ATENCIONAIS E ALARMES

Um pesquisador com diversas contribuições na modelagem de emoções em sistemas artificiais é Aaron Sloman, atualmente na Universidade de Birmingham – UK, onde dirige o projeto “Cognition and Affect” desde 1976 (entretanto em Birmingham desde 1991). Uma das peculiaridades que nos faz mencionar explicitamente a pesquisa desenvolvida por Sloman (2001) é seu modelo peculiar de emoções, bastante diferente dos modelos apresentados anteriormente.

Sloman (2001) desenvolveu seu modelo baseado na teoria da emoção como interrupção, desenvolvida anteriormente por Simon (1967). Nesta teoria, Simon também baseia-se na teoria de necessidades de Hull, mas com um resultado um pouco diferente. De acordo com Simon, existiriam três diferentes tipos de necessidades: aquelas oriundas de eventos ambientais incertos e não-preditivos (como, por exemplo, um predador que surge na floresta); as necessidades fisiológicas (como a fome, por exemplo), e as necessidades oriundas de associações cognitivas (como, por exemplo, as associações que podem levar à ansiedade). De acordo com Simon, even-

tos do ambiente, evocações da memória ou motivações poderiam alterar as metas correntes de um indivíduo, causando sua interrupção. A presença inesperada de um predador, por exemplo, poderia causar a interrupção do comportamento anterior de um animal e a geração de novas metas, como a tentativa de fuga do perigo. Simon ocupou-se de estudar profundamente o processo de interrupção de metas e a consequente mudança de comportamento causado pelas emoções. Sua teoria era a de que as emoções compreendiam um mecanismo de interrupção que funcionava como um monitor das diversas metas gerenciadas pelo sistema motivacional, causando mudanças dependentes de uma situação.

Baseado na teoria das emoções como interrupções de Sloman e Croucher (1981) Sloman (1987, 1992), então, desenvolve a teoria da penetração por filtros atencionais (AFP - *Attention Filter Penetration Theory*) para explicar as emoções. Ao contrário da teoria da interrupção de Simon, que encerrava terminalmente (ou seja, abortava os processos) os processos motivacionais interrompidos, na AFP de Sloman, os filtros atencionais disparados pelas emoções apenas causavam uma disrupção momentânea do curso de busca por uma meta, podendo retornar a ela posteriormente, caso o evento causador da emoção (e da disrupção) tenham sido satisfatoriamente tratados. Dessa forma, as emoções funcionariam como filtros atencionais ou, como Sloman (1998) acaba por chamar posteriormente, de *alarmes*. A ideia de um alarme é a de um processo (normalmente rápido e eficiente) que pode detectar padrões específicos e desencadear uma reação em cadeia que determina mudanças radicais no comportamento, diante dessa detecção. A justificativa para esses alarmes é a de que os processos de geração de comportamento podem ser algumas vezes demorados. Em algumas situações, consideradas críticas, talvez o sistema não tenha o tempo necessário para operar dessa maneira⁸, sendo que uma ação rápida é necessária, mesmo que não seja uma ação ótima. No mundo animal isso pode acontecer, por exemplo, em uma situação de predação, ou na iminência de outro perigo letal. No ser humano, isso pode acontecer sempre que situações particularmente perigosas são detectadas ou, de modo análogo, quando o aproveitamento de oportunidades para um comportamento altamente proveitoso demanda uma ação rápida. Dessa forma, Sloman

⁸ Ou seja, em situações críticas não se pode desenvolver processos de tomada de decisão demorados, tendo-se que agir rapidamente para tirar o sistema da situação crítica.

(1998) concebe um sistema emocional como um sistema de alarmes global que pode funcionar como um filtro atencional tanto em processos reativos como deliberativos, abortando certos tipos de comportamento e enfatizando outros tipos de comportamento.

5 CAÑAMERO E AS MODIFICAÇÕES EM MOTIVAÇÕES

O modelo desenvolvido por Lola Cañamero (1997, 1998, 2000, 2001, 2003, 2005) se diferencia dos anteriores em diversos aspectos. Esta pesquisadora propõe um sistema emocional para agentes autônomos em que as emoções funcionam como modificadores para motivações, a partir da definição de hormônios sintéticos que afetam os processos homeostáticos correspondentes às necessidades das criaturas virtuais.

O sistema motivacional desenvolvido por Cañamero presume um conjunto de *variáveis corporais* que geram *drives* para comportamentos regulatórios, visando manter os valores dessas variáveis dentro de parâmetros determinados. As variáveis utilizadas por ela são: nível de adrenalina, pressão sanguínea, taxa de açúcar no sangue, nível de endorfinas, nível de energia, batimento cardíaco, dor, taxa respiratória, temperatura e volume vascular. Cada uma dessas variáveis corporais possui um valor de referência, que é o valor desejado para a variável, bem como seu valor atual. A diferença entre o valor atual e o valor desejado funciona como *drive* para incitar certos tipos de comportamento que, além da ação que causam no ambiente, atualizam os valores das próprias variáveis corporais, em um mecanismo de homeostase. Assim, as ações da criatura no mundo, além de seu resultado pragmático no ambiente, efetuam a regulação das variáveis corporais.

A cada instante, a partir da diferença entre os valores atuais e os valores de referência das diversas variáveis corporais, define-se um *landscape* de *drives* (ou seja, um conjunto de índices de desempenho que informam o quanto as variáveis corporais estão distantes de seus valores de referência). O *drive* que tiver a maior intensidade (ou seja, que indicar uma variável corporal mais distante de seu valor de referência) é selecionado para determinar o comportamento imediato da criatura, gerando uma ação que atua sobre o ambiente e, ao mesmo tempo, reduz (lembrando que o *drive* é um número que indica o quanto uma determinada variável corporal está

distante de seu valor de referência) o *drive* selecionado. Esse é o sistema motivacional básico, ainda sem as emoções.

O mecanismo de emoções, segundo Cañamero, é independente do mecanismo motivacional, embora o afete. Determinadas situações do ambiente, que identificam situações críticas, disparam as emoções. As emoções utilizadas por Cañamero são: medo, raiva, felicidade, tristeza, enfado e curiosidade. Apenas uma emoção por vez está ativa. As situações do ambiente determinam diferentes níveis de *arousal*⁹ para cada uma das emoções, sendo que aquela com maior *arousal*, desde que este esteja acima de um limiar, é escolhida como a emoção corrente. A emoção irá gerar o disparo de um hormônio, que afetará as variáveis corporais, causando uma mudança em seus valores. O mesmo hormônio pode afetar variáveis corporais distintas. Alterando o valor das variáveis corporais, as emoções (ou seja, os hormônios) afetam o comportamento final da criatura, ao mexer com o *landscape* de *drives* que direcionam o comportamento da mesma.

No modelo de Cañamero, as emoções não estão diretamente ligadas ao caráter apraisivo das motivações, mas funcionam como modificadores de segunda ordem e/ou amplificadores de motivações, causando um desbalanço momentâneo do processo homeostático. Isso torna alguns *drives* mais “intensos” do que outros em situações especiais, permitindo que a criatura tenha um comportamento adequado para lidar com situações críticas, tais como situações que coloquem o indivíduo em situações de alto risco, ou situações extremamente favoráveis, que não se repetem com frequência, e que demandariam o aproveitamento de uma oportunidade ímpar que pode trazer benefícios expressivos aos objetivos da criatura.

6 PICARD E A COMPUTAÇÃO AFETIVA

Por fim, um último trabalho a destacarmos é o de Rosalynn Picard, não porque tenha contribuído com algum modelo específico de emoções, mas porque, de certa forma, estruturou a modelagem de emoções em sistemas artificiais em suas diversas manifestações. Picard (1997)

⁹ O termo “arousal” utilizado para representar o nível de intensidade, ou de atividade, ou estado intrínseco de excitação, é de difícil tradução, por expressar de maneira conjugada diferentes significados, e será mantido aqui em sua terminologia original, sem tradução, para evitar uma possível perda de conteúdo significativo.

organizou o estudo da área que chamou de Computação Afetiva, considerando os diferentes usos das emoções em sistemas artificiais. Em primeiro lugar, Picard chama a atenção que o comportamento de uma criatura artificial pode envolver emoções de três diferentes formas:

- Reconhecer Emoções
- Expressar Emoções
- Ter Emoções

Dessa forma, uma criatura (ou sistema) pode simplesmente ser capaz de reconhecer emoções em seres humanos e/ou outras criaturas, sem efetivamente expressá-las ou possuí-las de fato. Da mesma forma, ela pode ser capaz de expressar estados emocionais (simulá-los), sem, entretanto, utilizar cognitivamente esse estado para alterar seu comportamento. Isso pode acontecer, por exemplo, em criaturas que desejemos que cause um comportamento emocional no ser humano, sem que ela reaja emocionalmente. Por fim, uma criatura artificial pode realmente “ter” emoções, ou seja, ser afetada em seu comportamento por situações percebidas, de modo análogo ao que as emoções causam em um ser humano ou outro animal. É óbvio que, mesmo que essas habilidades sejam independentes, podemos ter o caso de um sistema artificial que, ao mesmo tempo, possa reconhecer emoções, expressar emoções e ter emoções.

Picard se propõe também a testar o “desempenho emocional” de um sistema. Nessa análise, ela se predispõe a avaliar o pretensão “comportamento emocional” de um sistema, tentando responder às seguintes perguntas sobre ele: O sistema parece ter emoções? Ele responde rapidamente a estímulos específicos? Ele interpreta a situação e a avalia? Ele parece possuir sentimentos diferentes para emoções diferentes? Essas emoções influenciam seu comportamento? Com relação ao efeito das emoções, ela indaga se as mesmas influenciam o aprendizado e/ou a tomada de decisões, ou se atua sobre a recuperação da memória.

Picard atenta para a possibilidade de haver diferentes níveis de representação para emoções. Em níveis mais baixos, podemos ter sinais emocionais (propriedades), tais como: o decaimento da resposta emotiva;

estímulos repetitivos; influências no temperamento e personalidade; não-linearidade; invariância no tempo; ativação; saturação; *feedbacks* cognitivo e físico; humor de background. No nível médio, analisa-se a questão da expressão das emoções, em termos de padrões e/ou modelos. Por fim, no alto nível, analisa-se o aproveitamento cognitivo das emoções, em termos dos conceitos e ideias e como estas são afetadas pelas emoções.

7 DISCUSSÃO

Os diversos trabalhos envolvendo emoções, tanto do ponto de vista puramente de modelagem cognitiva, como no tocante ao seu uso na construção de criaturas artificiais, possibilita levantar um grande número de considerações.

Em primeiro lugar, é necessário reconhecer que não existe um único modelo, nem um modelo que poderíamos identificar como sendo o *mainstream* na pesquisa sobre emoções. Por um lado, temos os estudos originários de Darwin sobre a expressão das emoções e o efeito que estas causam no corpo, o que é reconhecido por William James, Damásio e outros. A ideia de que as emoções afetam inicialmente o corpo, e a partir dele são incorporadas cognitivamente, nos leva à dicotomia *Appraisal/Arousal*¹⁰. O *Appraisal* diz respeito ao aspecto valorativo/avaliativo das emoções – uma medida da satisfação das necessidades intrínsecas da criatura (seja ela humana ou artificial), o quanto os propósitos da criatura foram ou não satisfeitos. Já o *Arousal* refere-se ao aspecto dinâmico das emoções – as reações específicas a objetos, eventos ou situações apontadas no modelo OCC.

A ideia de que o caráter avaliativo exposto acima diz respeito a um processo homeostático, onde certas variáveis internas devem ser reguladas, como em um sistema de controle, nos leva à influência das emoções no comportamento, funcionando como motivações para o ser humano, de modo que este seja capaz de tomar decisões. Entretanto, temos, aqui, a distinção do trabalho de Sloman, que enxerga as emoções somente em comportamentos de exceção, funcionando como mecanismos de alarmes ou mesmo como os modificadores ou amplificadores de valorações que afetam o julgamento, como apontado por Cañamero (2005). Vemos, por-

¹⁰ Como já levantado anteriormente, mantemos os termos originais sem tradução por entender que a tradução desses termos para o português poderia implicar uma possível perda de significado.

tanto, que o entendimento do que são realmente emoções se coloca como fragmentado e aparentemente cada um dos proponentes tem boas justificativas para embasar seus modelos. Entretanto, os modelos parecem ser radicalmente diferentes entre si. Será que poderíamos encontrar um modelo que unifique todos esses pontos, de tal forma que cada um dos argumentos possa ser incorporado?

Parece-nos que a grande dificuldade em se obter um modelo unificado é que existe um termo único: *emoção*, para o qual pretende-se que seja aplicado a diferentes partes de um processo de determinação do comportamento, onde cada autor parece enfatizar um aspecto diferente desse processo e identificá-lo com o termo *emoção*. Talvez essa dificuldade venha do uso não-técnico da palavra *emoção*, encontrado na linguagem comum, ou até mesmo no emprego de metáforas para lastrear o significado de uma palavra que não possui uma definição comum na linguagem cotidiana.

Nesse sentido, ouvindo os argumentos dos diferentes autores, e na tentativa de unificar as propostas, arriscaremos uma nova definição para emoções. Apesar de ser “mais uma” definição, em um tópico que já se encontra sobrecarregado delas, em nossa opinião é uma tentativa de unificar as diferentes visões, e talvez solucionar a questão da unificação, contemplando os diversos pontos de vista.

No modelo de sistemas motivacionais/emocionais que propomos neste artigo, entendemos que emoções podem ser vistas como “distorções cognitivas”¹¹, por vezes momentâneas, perdurando enquanto perdurem as situações, eventos ou objetos que as causam, as vezes mais perenes, como nos “estados de espírito” (moods) ou temperamento/personalidade, que visam aumentar a eficiência comportamental das criaturas em situações críticas.

¹¹ Tal fenômeno também poderia ser denominado “modulação” cognitiva, para talvez evitar uma possível interpretação pejorativa de que as emoções atrapalhariam a racionalidade, justamente o contrário do que queremos defender. Nossa perspectiva é que as emoções funcionam da mesma forma que aqueles espelhos existentes em parques de diversão, que deformam as pessoas, ou óculos que tingem de cores diferentes a realidade experimentada por meio deles, realçando aspectos dessa realidade que, as vezes, poderiam passar despercebidos. Algo como os cartunistas fazem ao retratar pessoas ou situações, enfatizando de maneira explícita certos aspectos distintivos da situação. Ao contrário de atrapalhar a racionalidade, ao enfatizar certos aspectos que merecem uma especial atenção da criatura, essa distorção visa evitar que a oportunidade de uma ação inteligente seja perdida. De certa forma, as emoções funcionariam como um sistema de “realidade aumentada”, em computação, tentando destacar elementos importantes da realidade que demandariam uma consideração especial.

Para defender essa proposta, nos serviremos de uma análise evolutiva sobre emoções em animais e o caráter evolutivo de seus cérebros. É notável que o comportamento motivacional existe inclusive em vertebrados evolutivamente mais antigos, tais como os peixes e répteis (até mesmo em alguns tipos de invertebrados como insetos). Apesar de uma potencial polêmica¹², o comportamento emocional é muito mais facilmente reconhecível em aves e mamíferos, onde o assim chamado sistema límbico é muito mais desenvolvido. Somente no ser humano (e talvez em alguns tipos de animais como macacos, chimpanzés, golfinhos e talvez outros), entretanto, é que o neo-córtex se encontra mais desenvolvido. Dessa forma, como aponta Cañamero, talvez os sistemas motivacionais tenham surgido inicialmente em espécies mais antigas, ainda sem capacidade emocional. O sistema emocional talvez tenha se desenvolvido posteriormente, a partir do surgimento do sistema límbico, conforme Panksepp (1982). Somente depois, com o desenvolvimento do neo-córtex, é que a razão foi incorporada. Apesar do surgimento de novos sistemas, eles não substituem os mais antigos, mas incorporam-se a eles e modificam seu funcionamento, agregando novas funcionalidades.

Nosso modelo aquiesce a proposta de Hull de que as criaturas (naturais e artificiais) devem possuir um certo repertório de “necessidades” e que uma medida da satisfação dessas necessidades é aferida em certos estados corpóreos, que identificaremos aqui como sendo “*drives*”. Eles podem fazer parte de sistemas motivacionais que determinam comportamentos reativos automáticos, desenvolvidos geneticamente pela evolução, levando a um processo de homeostase onde esses *drives* são reduzidos. Esse é o mecanismo que chamaremos de sistema motivacional básico, talvez encontrado em répteis e animais mais antigos na escala evolutiva (e que podem também ser implementados em criaturas artificiais).

Os sistemas motivacionais já são capazes de gerar um comportamento bastante sofisticado, uma vez que o processo de homeostase nos leva à satisfação das necessidades. Assim, cria-se um repertório condizente de necessidades, podemos desenvolver criaturas artificiais com um comportamento bastante complexo, capazes de agir de forma a atender a uma variada gama de propósitos. Observem que em nosso modelo, as características

¹² Algumas pessoas podem entender que animais sem sistema límbico seriam capazes de comportamento emocional

de *appraisal/arousal* não estão ainda ligados ao aspecto emocional, mas tão somente ao caráter motivacional do comportamento.

Sobre esse sistema motivacional (originalmente sem comportamento emocional), podemos adicionar essa ideia de uma “distorção cognitiva” sobre esse *landscape* de *drives*, alterando o comportamento motivacional em função de situações do ambiente. Qual seria a função dessa distorção cognitiva? Evolutivamente, essa é nossa hipótese, a natureza detectou que a identificação de situações que podem ser críticas para a sobrevivência de um agente (ou a identificação de situações excepcionalmente adequadas que não acontecem com frequência e que demandariam ser aproveitadas) poderia determinar ações urgentes e imediatas, que não poderiam ser postergadas em prol de ações menos urgentes. Dessa forma, o processo emocional teria surgido para “forçar” o comportamento das criaturas nessas situações críticas ao comportamento correto, do ponto de vista evolutivo, ou seja, aquele que maximiza as chances de sobrevivência da criatura e sua procriação, levando adiante seus genes.

Vamos analisar mais aprofundadamente essa hipótese da “distorção cognitiva” e sua relação com as diversas abordagens utilizadas nos diferentes modelos de emoções. De uma certa forma, os aspectos de *appraisal/arousal* estão aí presentes, uma vez que, alterando o espectro de *drives* da criatura, a avaliação da situação é modificada, bem como a possível intensidade dessa modificação. Da mesma forma, a hipótese de uma “distorção cognitiva” também contempla a ideia de Sloman de um alarme, ou um processo de atenção seletiva para essas situações que a determinam, uma vez que, ao serem disparadas, as distorções cognitivas funcionariam de certa forma como filtros atencionais, alarmes que determinam uma diferenciação do comportamento da criatura. Por fim, contempla-se, também, a ideia de Cañamero de um modificador para os *drives*, uma vez que esse processo de distorção cognitiva é implementado, de fato, na forma de modificadores sobre os diferentes *drives*, de forma análoga à proposta de Cañamero.

Defendemos, portanto, que a interpretação do mecanismo emocional como implementando uma “distorção cognitiva” momentânea ou perene, alterando o espectro de *drives* do sistema, e forçando o mesmo a um comportamento de exceção, apropriado para lidar tanto com situações críticas onde a própria existência do mesmo está em risco, ou o aprovei-

tamento de oportunidades excepcionais se mostra mandatário, é condizente com aproximadamente todos os modelos de emoções apresentados anteriormente. Assim, quiçá possa ser considerada como uma proposta unificadora dentre as teorias de emoções.

Um último ponto que seria importante mencionar a respeito do modelo diz respeito à expressão das emoções. De um ponto de vista puramente evolutivo, considerando-se a função cognitiva das emoções, como um modulador de motivações que age em situações críticas, distorcendo cognitivamente a importância relativa das motivações (de tal forma a evitar situações críticas ou aproveitar situações raras que podem trazer grandes benefícios), não haveria necessidade que essas mesmas emoções fossem expressas na forma de variáveis corporais como, de fato, acontece. Talvez exista uma segunda função cognitiva das emoções, relacionada à expressão das mesmas e com uma função social intra e interespecífica. Parece-nos que, talvez, a expressão das emoções tenha alguma função no estabelecimento de algum tipo de cooperação entre indivíduos (ou até mesmo no processo de competição), onde o conhecimento do estado emocional de outros indivíduos possa permitir o desenvolvimento de funcionalidades sociais que seriam mais difíceis caso essa emoção não fosse devidamente expressa e detectável por parte de outros indivíduos. Isso nos conecta, de certa forma, com a perspectiva social-construtivista das emoções, apontada por Cornelius (2000). Até onde temos conhecimento, esse uso da expressão das emoções para obtenção de comportamento social ainda não foi devidamente explorado em experimentos com criaturas artificiais. Existe um bom potencial para pesquisas nesse tópico, investigando essa possibilidade.

CONSIDERAÇÕES FINAIS

Como vimos ao longo deste capítulo, existem diversos modelos de emoções, que podem ser aproveitados em experimentos com criaturas artificiais. Apesar de aparentemente conflitantes, estes modelos talvez possam ser unificados em uma proposta comum, apresentada neste trabalho, onde as emoções são vistas como “distorções cognitivas”, disparadas ao se detectar situações críticas ou situações raras onde uma oportunidade não deve ser desperdiçada. Analisamos o que há em comum entre motivações e emoções e concluímos que o sistema emocional está profundamente atre-

lado ao sistema motivacional, mas que poderia ser distinguido deste. Dessa forma, apesar das emoções poderem direcionar o comportamento, não o fazem de maneira direta, mas indiretamente, modificando o espectro de motivações em função de condições especiais detectadas no ambiente.

Desta forma, propomos que seria mais produtivo, ao invés de tentarmos criar um “modelo de emoções”, que desenvolvêssemos modelos separados de sistemas motivacionais e emocionais, onde diferentes conceitos pudessem ser exercitados do ponto de vista funcional, e mesmo na impossibilidade de que pudéssemos identificar qual de suas partes deveria apropriadamente ser chamada de emoção, os fenômenos equivalentes encontrados em seres humanos e animais pudessem ser compreendidos e, independente da terminologia utilizada, pudessem ser utilizados na construção de criaturas artificiais.

Finalmente, um ponto que demanda maior investigação diz respeito ao estatuto ontológico que possivelmente poderia se atribuir a criaturas artificiais implementando o modelo que propomos aqui. Será que poderíamos afirmar que tais criaturas realmente possuem emoções, sentimentos, motivações, da mesma forma como criaturas naturais? Tais considerações extrapolam a perspectiva mais técnica que se buscou no presente trabalho, e demandam uma avaliação filosófica mais profunda que este autor deixa para os colegas filósofos, como um possível trabalho futuro. Deixamos aqui somente a pergunta, talvez um pouco incômoda: Por que não?

REFERÊNCIAS

- ARNOLD, M. B. *Emotion and personality*. New York, USA: Columbia University Press, 1960.
- BALKENIUS, C. Natural intelligence in artificial creatures. *Lund University Cognitive Studies*, v. 37, p. 286, 1995.
- BARRET, L. F. Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition and Emotion*, London, v. 12, n. 4, p. 579-599, 1998.
- BATES, J. The role of emotion in believable agents. *Communications of ACM*, New York, v. 37, n. 7, p. 122-125, July 1994.

BLUMBERG, B. M. *Old tricks, new dogs: ethology and interactive creatures*. 1997. Thesis (Doctorate of philosophy) – School Of Architecture And Planning, MIT, Cambridge, 1996.

BUCK, R. *Human motivation and emotion*. New York: Wiley, 1988.

CABANAC, M. What is emotion? *Behavioural processes*, Amsterdam, v. 60, n. 2, p. 69-83, Nov./Dec. 2002.

CAÑAMERO, L. A hormonal model of emotions for behavior control. *In: EUROPEAN CONFERENCE ON ARTIFICIAL LIFE*, 4., 1997. *Proceedings* [...]. Brighton: University of Sussex, 1997. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.50&rep=rep1&type=pdf>. Acesso em: 02 abr. 2007.

CAÑAMERO, L. Issues in the design of emotional agents. *In: FALL SYMPOSIUM*, 1998, Menlo Park. *Technical report...* Menlo Park: AAAI Press, 1998. p. 49-54.

CAÑAMERO, L. Designing emotions for activity selection. *DAIMI Report Series*, Katrinebjerg, Denmark, v. 29, n. 545, Feb. 2000.

CAÑAMERO, L. Emotions and adaptation in autonomous agents: a design perspective. *Cybernetics and Systems: An International Journal*, New York, v. 32, n. 5, p. 507-529, 2001.

CAÑAMERO, L. Designing emotions for activity selection in autonomous agents. *In: TRAPPL, R.; PETTA, P.; PAYR, S. (ed). Emotions in humans and artifacts*. Cambridge: MIT Press, 2003.

CAÑAMERO, L. Emotion understanding from the perspective of autonomous robots research. *Neural networks*, New York, v. 18, n. 4, p. 445-455, May 2005.

CARRUTHERS, G. Types of body representation and the sense of embodiment. *Consciousness and cognition*, Maryland Heights, v. 17, n. 4, p. 1302-1316, Dec. 2008.

CORNELIUS, R. R. Theoretical approaches to emotion. *In: ISCA TUTORIAL AND RESEARCH WORKSHOP ON SPEECH AND EMOTION*, 2000, Newcastle. *Proceedings* [...]. Newcastle: ISCA, 2000.

DAMÁSIO, A. R. *O erro de descartes: emoção, razão e cérebro humano*. Sintra: Publicações Europa-América, 1994.

DAMÁSIO, A. R. *The feeling of what happens: body and emotion in the making of consciousness*. New York: Harcourt, 1999.

- DAMÁSIO, A. R. *Looking for Spinoza: joy, sorrow and the feeling brain*. London: William Heinemann, 2003.
- DARWIN, C. *The expression of the emotions in man and animals*. Chicago: University of Chicago Press, 1965.
- DEAN, J. Animats and what they can tell us. *Trends in Cognitive Sciences*, Oxford, v. 2, n. 2, p. 60-67, Feb. 1998.
- DIXON, T. *From passions to emotions: the creation of a secular psychological category*. New York, US: Cambridge University Press, 2003.
- DIXON, T. *Emotion: The history of a keyword in crisis*. *Emotion Review*, London, v. 4, n. 4, p. 338-344, Aug. 2012.
- EL-NASR, M. S. *Modeling emotion dynamics in intelligent agents*. 1998. Thesis (Master's of Computer Science) – Dwight Look College of Engineering, Texas A e M University, Texas, 1998.
- EL-NASR, M. S.; YEN, J; IOERGER, T. R. FLAME – Fuzzy logic adaptive model of emotions. *Autonomous agents and multi-agent systems*, New York, v. 3, n. 3, p. 219-257, Sep. 2000.
- FRANKLIN, S. *Artificial Minds*. Cambridge, MA: Bradford Books; MIT Press, 1995.
- GOLEMAN, D. *Inteligência Emocional*. Rio de Janeiro: Objetiva, 1995.
- HULL, C. L. *A Behavior System*. New Haven: Yale University Press, 1952.
- IZARD, C. E. The many meanings/aspects of emotion: definitions, functions, activation, and regulation. *Emotion Review*, London, v. 2, n. 4, p. 363-370, Oct. 2010.
- JAMES, S. *Passion and action: the emotions in seventeenth-century philosophy*. Oxford: Clarendon Press, 1997.
- JAMES, W. What is an emotion? *Mind*, Oxford, v. 9, n. 34, p. 188-205, Apr. 1884.
- JOHNSON, G. *Theories of emotion*. In: INTERNET Encyclopaedia of Philosophy (IEP): a peer-reviewed academic resource, 2009. Disponível em: <https://www.iep.utm.edu/emotion/>. Acesso em: 25 de maio de 2017.
- KODA, T.; MAES, P. Agents with faces: the effect of personification. In: INTERNATIONAL WORKSHOP ON ROBOT AND HUMAN COMMUNICATION RO-MAN'96 TSUKUBA, 5., 1996, Japan. *Proceedings* [...]. Japan: IEEE, 1996. p. 189-194.

- LANGLEY, P; LAIRD, J. E; ROGERS, S. Cognitive architectures: research issues and challenges. *Cognitive Systems Research*, New York, v. 10, n. 2, p. 141-160, June 2009.
- MALFAZ, M.; SALICHS, M. A. A new architecture for autonomous robots based on emotions. *IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, Lisbon, v. 37, n. 8, p. 805-809, July 2004.
- MASLOW, A. H. A theory of human motivation. *Psychological review*, Washington, v. 50, n. 4, p. 370-396, 1943.
- MEHRABIAN, A. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology*, New York, v. 14, n. 4, p. 261-292, Dec. 1996.
- MEYER, J. J. Ch. Reasoning about emotional agents. *International Journal of Intelligent Systems*, Hoboken, v. 21, n. 6, p. 601-619, Apr. 2006.
- MOORS, A.; ELLSWORTH, P. C.; SCHERER, K. R.; FRIJDA, N. H. Appraisal theories of emotion: state of the art and future development. *Emotion Review*, London, v. 5, n. 2, p. 119-124, Mar. 2013.
- OKA, N. *et al.* Embodiment without a physical body. In: ANNUAL MEETING OF THE COGNITIVE SCIENCE SOCIETY, 23., 2001. *Proceedings [...]*. Edinburgh, 2001. p. 48-52.
- ORTONY, A. On making believable emotional agents believable. In: TRAPPL R. *et al.* (ed.). *Emotions in Humans and Artifacts*. Cambridge, MA; London, UK: MIT Press, 2003. p. 189-212.
- ORTONY, A.; CLORE, G.; COLLINS, A. *The cognitive structure of emotions*. Cambridge: Cambridge University Press, 1988.
- PANKSEPP, J. Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, Cambridge, v. 5, n. 3, p. 407-422, Sept. 1982.
- PICARD, R. W. *Affective Computing*. Cambridge, MA: MIT Press, 1997.
- POSNER, J., RUSSELL, J. A; PETERSON, B. S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, Cambridge, v. 17, n. 3, p. 715-734, Sept. 2005.
- REILLY, W. S. N. Believable social and emotional agents. 1996. Thesis (Doctorate of philosophy) – School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996.

- REISENZEIN, R. Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology*, Washington, v. 67, n. 3, p. 525-539, 1994.
- RORTY, A. O. From passions to emotions and sentiments. *Philosophy*, Cambridge, v. 57, n. 220, p. 159-172, Jan. 1982.
- RUSSELL, J. A. A circumplex model of affect. *Journal of Personality and Social Psychology*, v. 39, n. 6, p. 1161-1178. 1980.
- RUSSELL, J. A; MEHRABIAN, A. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, Maryland Heights, v. 11, n. 3, p. 273-294, Sep. 1977.
- SARMENTO, L. M. *An emotion-based agent architecture*. 2004. Tese (Mestrado em Inteligência Artificial e Computação) – Faculdade de Ciências, Universidade do Porto, Porto, 2004.
- SCHLOSBERG, H. Three dimensions of emotion. *Psychological review*, Washington, v. 61, n. 2, p. 81-88, Mar. 1954.
- SIMON, H. A. Motivational and emotional controls of cognition. *Psychological review*, Washington, v. 74, n. 1, p. 29-39, Jan. 1967.
- SLOMAN, A. Motives, mechanisms and emotions, emotion and cognition. *In*: BODEN, M. A. (ed.). *The Philosophy of Artificial Intelligence “Oxford Readings in Philosophy”*. Oxford: Oxford University Press, 1987. p. 217-247. (Series Oxford University Press).
- SLOMAN, A. Prolegomena to a theory of communication and affect. *In*: ORTONY, A.; SLACK, J.; STOCK, O. (ed.). *Communication from an artificial intelligence perspective: theoretical and applied issues*. Heidelberg, DE: Springer, 1992. p. 229-260. Available in: <http://www.cs.bham.ac.uk/research/cogaff>. Acesso in: 12 may 2012.
- SLOMAN, A. Damasio, descartes, alarms, and metamangement. *In*: INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS, 1998, San Diego. *Proceedings*[...]. San Diego: IEEE, 1998. p. 2652–2657. Disponível: <http://citeseer.ist.psu.edu/cachedpage/65292/1>. Acesso em: 09 abr. 2007.
- SLOMAN, A. Beyond shallow models of emotion. *Cognitive Processing*, Heidelberg, v. 2, n. 1, p. 177-198. 2001. Disponível em: <http://www.cs.umbc.edu/courses/graduate/691m/spring02/papers/sloman-emotion.pdf>. Acesso em: 09 abr. 2007.

- SLOMAN, A.; CROUCHER, M. Why robots will have emotions. *In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 1981, Vancouver. *Proceedings*[...]. Vancouver: University of British Columbia, 1981. v. 1. p. 197-202.
- STEELS, L. The artificial life roots of artificial intelligence. *Artificial life*, Cambridge, v. 1, n. 1/2, p.75-110, Oct. 1993.
- TOMLINSON, B.; BLUMBERG, B. Social synthetic characters. *ACM SIGGRAPH Computer Graphics*, New York, v. 36, n. 2, p. 5-7, May 2002. doi:<http://dx.doi.org/10.1145/566656.566659>.
- VELÁZQUEZ, J. A computational framework for emotion-based control. *In: GROUNDING EMOTIONS IN ADAPTIVE SYSTEMS WORKSHOP*, Zurich, 1998. *Proceedings* [...]. Zurich: Austrian Research Institute for Artificial Intelligence, 1998.
- VELÁZQUEZ, J. D. From affect programs to higher cognitive emotions: an emotion-based control approach. *In: EMOTION-BASED AGENT ARCHITECTURE WORKSHOP AT THE INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS*, 1999. *Proceedings*[...]. Seattle, 1999. p. 114-120.
- VENTURA, R. *Emotion-based agents*. 2000. 59 f. Tese (Mestrado em Engenharia Electrotécnica e de Computadores) – Instituto Superior Técnico, Lisboa, 2000.
- WALLACH, W.; ALLEN, C. *Moral machines: teaching robots right from wrong*. Oxford: Oxford University Press, 2008.
- WILSON, M. Six views of embodied cognition. *Psychonomic bulletin and review*, New York, v. 9, n. 4, p. 625-636, Dec. 2002.
- WOOLDRIDGE, M; JENNINGS, N. R. *Agent theories, architectures, and languages: a survey*. Berlin: Springer-Verlag, 1995.
- ZIEMKE, T; LOWE, R. On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognitive computation*, New York, v. 1, n.1, p. 104-117, Mar. 2009.