Universidade Federal do Rio Grande do Sul
Centro Interdisciplinar de Novas Tecnologias na Educação
Programa de Pós-Graduação em Informática na Educação

# AGÊNCIAS DO ARTIFICIAL E DO HUMANO:

## uma análise de noções do humano na Inteligência Artificial a partir de perspectivas sociais e culturais

Rafael Wild

Porto Alegre
2011

Rafael Wild

# AGÊNCIAS DO ARTIFICIAL E DO HUMANO:
## uma análise de noções do humano na Inteligência Artificial a partir de perspectivas sociais e culturais

Tese apresentada como requisito parcial para obtenção do título de doutor junto ao Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul.

Orientadora: Profa. Dra. Maria Cristina Villanova Biazus

Co-Orientadora: Profa. Dra. Cleci Maraschin

Porto Alegre

2011

CIP – Catalogação na Publicação

Wild, Rafael

Agências do artificial e do humano: uma análise de noções do humano na Inteligência Artificial a partir de perspectivas sociais e culturais. / Rafael Wild. – 2011.

160 f.

Orientadora: Maria Cristina Villanova Biazus.

Co-orientadora: Cleci Maraschin.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul, Centro Interdisciplinar de Novas Tecnologias na Educação. Programa de Pós- Graduação em Informática na Educação, Porto Alegre, BR-RS, 2011.

1. Inteligência Artificial. 2. Ciência, Tecnologia e Sociedade. 3. Humano. 4. Emoção. 5. Conhecimento. I. Biazus, Maria Cristina Villanova, orient. II. Maraschin, Cleci, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo autor.

Rafael Wild

# AGÊNCIAS DO ARTIFICIAL E DO HUMANO:
## uma análise de noções do humano na Inteligência Artificial a partir de perspectivas sociais e culturais

Tese apresentada como requisito parcial para obtenção do título de doutor junto ao Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul.

Aprovada em 2 de junho de 2011.

Orientadora: Profa. Dra. Maria Cristina Villanova Biazus

Co-Orientadora: Profa. Dra. Cleci Maraschin

Prof. Dr. Eliseo Berni Reategui - UFRGS

Prof. Dr. Arlei Sander Damo - UFRGS

Prof. Dr. Rafael Vetromille-Castro - UFPEL

À memória de Mário José Lopes Guimarães Júnior.

## Agradecimentos

Esta tese foi realizada com o suporte do Conselho Nacional de Pesquisa (CNPq) e da Coordenação de Apoio e Pesquisa em Ensino Superior (CAPES). O Programa de Pós-graduação em Informática na Educação, da Universidade Federal do Rio Grande do Sul, abrigou o desenvolvimento deste trabalho. Durante o trabalho de campo em Portugal, quem proporcionou o ambiente e a estrutura foi o Instituto de Engenharia e Sistemas Investigação e Desenvolvimento – INESC-ID.

Uma tese é, além de um texto acabado, também uma trajetória na vida intelectual. Agradeço a quem apontou e estimulou o início da caminhada: Patrícia Kirst, Magda Bercht, e, com carinho especial, Mário Guimarães.

Merecem uma menção particular as pessoas que apostaram nesta trajetória, que abriram caminhos e emprestaram uma bússola em vários momentos: minhas orientadoras, Maria Cristina Villanova Biazus e Cleci Maraschin, e também Rosa Vicari e Ana Paiva.

As reuniões instigantes e fraternas de nosso grupo de pesquisa, o NESTA, dispararam muitas das nossas melhores ideias.

Também agradeço aos colegas e amigos com quem aprendi sobre a vida acadêmica e fora dela: Paulo, Sílvia, Paka, Vanessa, Júlio, Daniela, Lourenço. Elaine e Erika, e Patrícia J., obrigado pela confiança. Marcelo Mercante, amigo que ganhei como dádiva. E à Paula, que com paciência ouviu minhas ideias desde 2007, e me ensinou muito sobre ser *humano*. Em Portugal, às minhas comunidades: Raquel (obrigado muito especial), Joa, Simone, Victor, e Pedro, Gonçalo, e Rui.

Este trabalho tem uma dívida com todos os participantes, cujos nomes ganharam versões criativas, e que dispuseram do seu tempo e sua atenção para que eu pudesse aprender com eles. Obrigado!

## Resumo

Esta tese analisa noções construídas sobre o ser humano que são apropriadas com fins tecnológicos, em sistemas computacionais produzidos por praticantes da Inteligência Artificial. Foi desenvolvido com base em um trabalho de campo de observação participante junto a grupos de pesquisa acadêmico na área de Inteligência Artificial, um brasileiro e outro europeu (português). O trabalho articula-se com as demandas da Informática na Educação ao focar, de maneira não estrita, projetos com caráter pedagógico. O presente estudo, através dos significados e as práticas observadas a partir de dentro dos grupos, procurou compreender o conhecimento do participante enquanto pertencente a uma cultura própria e peculiar, e a lógica interna desta cultura. Foram interrogados com especial atenção os artefatos produzidos: sistemas computacionais, investidos das características funcionais desejadas pelos participantes, e materializando suas práticas e premissas. Observou-se como emoção, conhecimento, cultura, e agência, entre outros, são conceituados, estabelecidos e colocados em práticas como categorias do humano, não apenas como definições expressas em texto, mas como materializadas em artefatos e em expectativas sobre o encontro entre estes artefatos e seus usuários. Foi consistentemente trabalhado o "colocar em perspectiva" das práticas e noções próprias do campo estudado, a partir de ferramentas teóricas propostas pelos Estudos de Ciência e Tecnologia, em especial por B. Latour, L. Suchman e D. Forsythe. As práticas e noções, no campo abordado, são conhecimento científico e tecnológico, com estatuto próprio e estabelecido como válido e legítimo; em relação a isto, foi sistematicamente buscada a colocação desta validade e desta legitimidade *em perspectiva*, mostrando como esta validade relaciona-se com a forma de produção e legitimação, e como esta produção e legitimação podem ser vistas de outras formas. Espera-se, com estes resultados, contribuir para um diálogo mais sofisticado dentro da Informática na Educação entre as práticas tecnológicas, a Ciência da Computação e Inteligência Artificial, e a aplicação social e pedagógica destas práticas.



Palavras-chave: Inteligência Artificial; Ciência, Tecnologia e Sociedade; Computação Afetiva; Humano; Conhecimento; Emoção.

## Abstract

This thesis addresses notions of human that are present in computer-based systems built by researchers in the area of Artificial Intelligence. Participant observation was performed in fieldwork within two academic research groups in Artificial Intelligence; one of such groups is Brazilian, while the other is Portuguese. The focus is on research projects displaying a pedagogical orientation. This thesis aims at understanding meanings and practices current in the groups, understood as local cultural settings, and the logics that underpin such meanings and practices. The technological artifacts that comprises their work, computer systems invested of certain functional characteristics, were interrogated. Categories such as emotion, knowledge, culture, and agency were followed as they are conceptualized and deployed as human traits, not only as textual definitions, but also as artefactual materializations and expectations about how users should encounter these artifacts. As a methodological analytics, these practices and notions were systematically compared with alternative perspectives, drawn from the theoretical references of the Science and Technological Studies (with special mention to B. Latour, L. Suchman and D. Forsythe). The validity and legitimacy of the positions of the group were not denied or devalued in this analytical process, but instead subjected to inquiry from different perspectives. The aims are making visible the relation of this validity and legitimacy with specific, situated processes of production and legitimation, and proposing that these processes could be considered in other, different ways.

Keywords: Artificial Intelligence; Science and Technology Studies; Affective Computing; Human; Emotion; Knowledge.

# Sumário

# 1 Introdução

O que ao certo quer dizer “ser humano” é uma questão que segue sendo colocada em nossa época com a mesma ansiedade que tem merecido ao longo da história. As respostas que formulamos hoje são múltiplas, e mudam conforme a perspectiva de quem responde, é claro. O presente trabalho tem seu interesse focado em uma destas perspectivas: a noção de humano que é mobilizada por um grupo específico de pessoas, os cientistas e engenheiros envolvidos com a Inteligência Artificial e a Computação Afetiva.

A ciência e tecnologia, como construção e processos de apropriação de saberes e artefatos, possuem um papel central na ideia que nossa sociedade faz de si. A tecnologia computacional, e a informática em especial, possuem uma participação muito profunda no processo de constituição de nossa sociedade contemporânea, ocupando espaços em práticas já estabelecidas e ao mesmo tempo sendo central para a invenção de novas práticas, em um sem-número de âmbitos do viver social. Apenas para situar a discussão, podemos citar alguns exemplos: a comunicação entre pessoas com correio eletrônico, mensagem "instantânea" e videotelefonia; a comunicação do cidadão com instituições e governo através de correio eletrônico e interação com página internet; a experiência de escolarização por ensino a distância e mediada por sistemas tutores computacionais; compras e propaganda na internet; e a experiência do corpo e da imagem do corpo nas formas de trabalho e de doenças funcionais associadas à computação, e em visualizações corporais de caráter médico (ecografia, ressonância). O saber tecnológico vinculado à computação e à informática possui um prestígio especial, uma vez que ao ser exercido efetivamente torna-se criador de espaços de atividade e vivência pelos quais circulam pessoas de todos os lugares sociais, processo de criação que é percebido como engendrado e tornado possível pela tecnologia.

Profissionais que trabalham com a computação e a informática constroem seus artefatos e sistemas para pessoas utilizarem: os chamados usuários. O uso proposto e regulado desta tecnologia é informado pela ideia que os profissionais deste campo têm das pessoas; ao mesmo tempo, a tecnologia vai encontrar-se com as pessoas e transformá-las, mudando sua forma de lidar com aquela prática em que houve o encontro com a tecnologia, ou mesmo criando novas práticas dentro de sua vida.

O encontro com a tecnologia não é linear, e sua apropriação pelos usuários é parte de um processo em que estão envolvidos, além dos usuários, os profissionais tecnológicos e outros interessados na existência e controle de um determinado artefato material ou técnica: os representantes dos interesses comerciais da organização que promove a tecnologia, usuários que desejam utilizá-la mas têm seu acesso impedido, pessoas com diversas posições políticas a respeito da forma como a tecnologia vai incidir sobre a prática que ela promove (P. Edwards, 2001; Pfaffenberger, 1988). Em particular, construir e colocar em uso um sistema computacional é, portanto, uma forma de configurar práticas e sujeitos sociais, e uma forma que é posta em uso com muita frequência em nossa sociedade (P. Edwards, 2001; Guimarães Jr., 2005). O artefato computacional, constituído pelo software, seu equipamento e infraestrutura material e modos pensados de uso, materializa uma noção de usuário que vai encontrar um grupo real de pessoas, e que neste processo vai gerar realizações, tensões e mudanças. Esta situação será visível nas práticas e nos modos de ser das pessoas enquanto usuárias, e também incidirá no próprio artefato e nas práticas dos profissionais e organizações responsáveis por ele (Guimarães Jr., 2005).

Alguns campos da ciência e da tecnologia, entretanto, chamam a atenção por dedicarem-se explicitamente a processos de constituição e definição do ser humano. Estes campos, como a Inteligência Artificial e a Biotecnologia, empenham-se ativamente em materializar, em corpos humanos e em artefatos que a eles se acoplam ou que os substituem, diálogos entre o que deve ser considerado *natural* e o que pode ser criado como *artificial*; entre o *futuro* que se empenham em imaginar e substancializar, e o *presente/passado* como referência a partir do qual progredir; entre as virtudes e fraquezas *humanas* e sua contrapartida, sua replicação e superação *tecnológica*.

As práticas que surgem vinculadas a estes campos científicos e tecnológicos têm transformado significativamente a ideia que as pessoas têm daquilo que *podem desejar* e

daquilo que *devem realizar*. A escala em que estes campos se articulam com a sociedade, mobilizando economias, colocando em xeque a fundamentação das leis, e transformando corpos e paisagens, faz com que entender como a perspectiva de humano que circula nestes universos tecnológicos seja de grande relevância. Longe de estabelecer fronteiras definidas e estatutos éticos claros a partir das dicotomias que a ciência e a tecnologia percebem, os diálogos acima referidos têm borrado as fronteiras entre seus polos e desconstruído parâmetros que estavam em uso para compreendê-los.

O futuro, neste contexto, é uma noção que age no *agora* sobre as formas como entendemos o humano. Mais do que simplesmente um tempo que ainda não ocorreu, é um tempo carregado, no presente, com narrativas de avanços tecnológicos e com existências consideradas como dadas, mais do que esperadas. Postulamos um futuro de artefatos *humanizados,* e humanos portando extensões e correções *artefatuais* de si. Este futuro informa a prática tecnológica e os desejos dos indivíduos *hoje*; a relação entre o humano e a máquina desenrola-se ao longo de nosso tempo, na definição adequada de Katherine Hayles (2005), como "um atrator estranho, definindo o espaço de fase dentro do qual trajetórias narrativas podem ser traçadas"[1].

Abordamos, na presente pesquisa, projetos em que busca-se replicar, através de tecnologia/computador, características ditas próprias do ser humano, como inteligência (inteligência artificial); emoções (computação afetiva); corporeidade (robótica humanoide); agência (agentes de software); sociabilidade e redes sociais. Projetos deste tipo compartilham como traço comum o interesse em replicar competências próprias do humano em artefatos tecnológicos. Grupos que envolvem engenheiros e cientistas de diversas áreas, tais como computação, psicologia e neurociências, buscam construir sistemas computacionais cujo diferencial, em relação a outras formas de objetos informáticos, é sua semelhança proposta com o humano no que se relaciona com seu pensar e agir. Estes grupos e seus participantes, dependendo do tipo de problema e de solução que investigam, filiam-se a áreas que denominam com nomes sugestivos: Inteligência Artificial, Computação Afetiva, Sistemas Tutores Inteligentes, Engenharia de Software Orientada a Agentes. O pertencimento a estas áreas não é mutuamente

1 "A strange attractor, defining the phase space within which narrative pathways may be traced". Traduções de citações apresentadas em nota de rodapé na língua original são minhas.

exclusivo, já que métodos, técnicas e premissas teóricas costumam ser compartilhados, herdados e modificados entre elas. Ao denominar sua área e filiar seu trabalho, no entanto, estes pesquisadores tornam visível sua abordagem particular da busca, mais geral, pela automatização das competências e sensibilidades humanas.

A Informática na Educação faz parte deste contexto, e a recriação de características humanas é procurada de forma explícita pelos seus praticantes como maneira de tornar a tecnologia educacional melhor e mais adequada para as pessoas envolvidas nos processos da educação (Porayska-Pomsta & Pain, 2004; Vicente & Pain, 2002). O universo desta pesquisa foi construído tendo estas considerações em vista, e constitui-se de dois grupos de pesquisa em Inteligência Artificial, um grupo brasileiro e um grupo europeu (português), cujos projetos são direcionados para ou muito próximos à Informática na Educação. Estes grupos são de caráter acadêmico, reunindo professores de universidade, alunos de pós-graduação e outros pesquisadores. Seus projetos colocam em ação diversas técnicas e áreas da IA, tais como agentes artificiais, computação afetiva, IA simbólica, e até robótica, com o objetivo de criar sistemas *inteligentes*, interativos em sua maioria, aplicados em apoio à educação ou em atividades pedagógicas em um sentido amplo. As noções mencionadas, e as formas como se fazem presentes no trabalho destas pessoas, foram buscadas durante o processo de pesquisa através da comparação entre as ideias apresentadas na literatura e pela produção artefatual considerada como referências nos grupos, da produção textual de pesquisadores na área, e da observação em campo dos pesquisadores e de seus artefatos. Este referencial de ideias e de artefatos foi analisado e colocado em perspectiva na comparação com outras formas de compreender o humano e suas práticas e saberes, presentes em disciplinas nas quais foram procuradas as referências teóricas utilizadas para este trabalho.

A produção destas áreas da IA e da tecnologia é expressa, pelos seus membros, como um processo em duas vias: como a construção de sistemas baseados em modelos do humano, e como o aperfeiçoamento destes modelos utilizando os sistemas construídos como ferramentas. A Inteligência Artificial propõe-se, assim, a *criar máquinas e computadores* que raciocinam ou que realizam atividades associadas à inteligência humana, e a *estudar, através da computação*, o pensamento e o comportamento humanos (Russell & Norvig, 1995, p. 5). A Computação Afetiva, por sua vez, propõe-se a estender o objeto de estudo clássico da Inteligência Artificial, a inteligência, e *transferir para*

*sistemas computacionais* competências emocionais e sociais (Porayska-Pomsta & Pain, 2004). Como o artífice que emprega suas habilidades para construir um objeto que se assemelhe ou que funcione *como* a obra original, engenheiros e cientistas destas áreas trabalham para construir *réplicas* nas quais possam ver as características pelas quais eles reconhecem o humano original. Esta é uma situação contemporânea que ocorre em um contexto histórico, mas, como vimos, também dentro de uma tradição que imagina e deseja o futuro de nossa sociedade através de suas criações tecnológicas.

Tendo presente esta situação, propomos pensar neste trabalho as *características*, entre as quais citamos a inteligência, a emoção e a sociabilidade, que os *praticantes* da Inteligência Artificial e Computação Afetiva colocam em foco e procuram *replicar,* e que por cujo meio reconhecem o *humano*. O objetivo será pensar como estas características são construídas pelos praticantes dentro dos seus processos de criação de sistemas computacionais, e explorar como, por meio da formulação e utilização destas características, o humano é figurado e construído por estas pessoas. Nossa intenção não se resume a re-elencar uma série de categorias propostas alhures: procuramos, sim, preenchê-las com os significados que tomam ao serem expressas, discursivas ou materialmente, no trabalho destes atores. Nossa visão é a de que os significados atribuídos ao "ser humano" são continuamente reconfigurados também dentro da área tecnológica, e que por vezes a utilização de categorias consideradas naturais passa ao largo de considerações sobre sua construção cultural e sua contingência histórica, isto é, sobre como tornam-se naturais por um efeito de *naturalização*. Por este motivo, é importante repensar estas categorias e o próprio processo de atribuição de sentidos a elas. Realizado com a utilização de métodos baseados em observação participante e análise etnográfica, procuramos compreender e elucidar, a partir das observações em campo e na produção acadêmica dos grupos, o humano pensado e agenciado por seus participantes.

Acreditamos, em suma, poder contribuir, através da produção de conhecimento que seja acessível aos agentes do campo, com a possibilidade de realimentar as práticas tecnológicas pesquisadas com diretrizes sociais e culturais, e com a construção de novas perspectivas de abordagem destas práticas, tanto em relação à análise de projetos como em relação à sua concepção.

## *1.1 Produção de tecnologia e de sociedade: contribuindo para um debate*

Cientistas e programadores, ao falarem sobre a produção de sistemas de computação com características humanas, consideram a questão partindo de formas de pensar o humano que constituem-se em traços corriqueiros do campo científico-tecnológico em que estão imersos. Como um exemplo destas formas de pensar, podemos citar a formulação de atributos intrínsecos de componentes humanos e computacionais dos sistemas como caracteres naturais que podem ser mensurados e replicados individualmente em cada componente. No entanto, estas não são as únicas maneiras de entender o humano e seu agir subjetivo e social.

É importante observar com atenção estas premissas e conceitos e entendê-las como contingentes, isto é, não universais, concepções sobre o humano próprias da visão de mundo da tecnologia e por isso bastante específicas. Também é importante entendê-las como emergentes de relações sociais, consolidadas a partir de noções sobre o que é o humano e suas condutas próprias enquanto tal, noções estas que refletem o ponto de vista que os produtores de tecnologia computacional têm do humano.

O que propomos neste trabalho é contribuir para uma perspectiva mais ampla do que é tecnologia e do que é humano, uma perspectiva que valorize compreender as apropriações e interpretações das técnicas e dos objetos e que, em constante processo, geram novos e diferentes fatos culturais em novas circunstâncias de uso: outras realidades sociais, outras formas de ver o mundo, pessoas utilizando os artefatos não de forma isolada mas fazendo parte de um contexto maior. Aqui interessa partir do *fato tecnológico* original do sistema computacional desenvolvido por cientistas em um laboratório para usuários "modelo" e abrir a possibilidade de *sentidos e fazeres* tecnológicos a par de sua construção mútua com os sentidos e fazeres da sociedade e sua diversidade.

A Informática na Educação não possui uma formação disciplinar estável: é uma área de estudo e aplicação recente, e congrega uma variedade de concepções sobre o que constitui seu objeto de estudo, quais devem ser seus objetivos e que paradigmas metodológicos são mais adequados. Estes aspectos mencionados, para a Informática na Educação, não são *indefinidos*, e sim múltiplos na medida em que esta se caracteriza como

um *campo* interdisciplinar de estudos (e não uma disciplina). Dentro deste campo de estudos, uma das contribuições importantes estabelecidas é a mobilização das tecnologias digitais e computacionais, que tornaram-se muito importantes para nossa sociedade, para prover novas possibilidades educativas e aperfeiçoar processos educacionais existentes, assim como para enriquecer a formação dos educandos em tópicos e competências relacionados a estas tecnologias educacionais.

Seguindo esta linha contributiva, a Inteligência Artificial ganhou um papel importante dentro da Informática na Educação, em função da própria natureza dos problemas que examina, dos recursos e temas que agrega, e do tipo de sistemas que produz. É presente em projetos de Sistemas Tutores Inteligentes, de Ambientes Virtuais de Ensino, de narrativa emergente, de agentes pedagógicos animados, e possui sua própria trilha no mais importante simpósio brasileiro de Informática na Educação. E, por fim, constitui um grupo ativo de pesquisa e produção dentro do próprio programa de pós-graduação em que esta tese se desenvolve.

Dada esta presença e esta importância da IA dentro da Informática na Educação, compreender e dialogar com suas premissas e noções relacionadas ao humano, de maneira reflexiva e a partir de outras perspectivas – que por sua vez estão presentes em outras abordagens da Informática na Educação – é uma rota promissora. A intenção é contribuir para enriquecer o próprio diálogo da IA e da Informática na Educação, de seus praticantes e de seus artefatos, com o amplo universo humano.

Incorporando-se esta perspectiva, espera-se contribuir com subsídios para propostas de intervenção na realidade social baseadas ou mediadas pelas tecnologias em questão. Tais subsídios podem futuramente operar, por exemplo, na avaliação de tecnologias e sistemas ou como requisitos de projeto tecnológico, incorporando considerações como as discutidas neste trabalho desde a própria concepção do artefato. Esta pesquisa insere-se no panorama contemporâneo da Informática na Educação na medida em que discussões tais como a da presente proposta podem auxiliar na orientação de agentes ligados à produção tecnológica - pesquisadores acadêmicos, agentes da indústria, agentes governamentais responsáveis pelas políticas públicas relacionadas com o tema - principalmente naqueles projetos que interfaceiam diretamente com sujeitos "não-tecnológicos", tais como sistemas de apoio a ensino. Em relação à importância acadêmica do trabalho, destaca-se ainda a

relevância de utilizar uma forma de análise já estabelecida - pensar a natureza social da tecnologia - aplicando-a a um caso novo, principalmente tratando-se de Brasil. Consideramos que este é um caso privilegiado de estudo, tendo em vista que os resultados promissores da tecnologia educacional e o prestígio associado à recriação de características humanas favorecem a naturalização de pressupostos e concepções do que pode ser considerado como “humano”.

# 2 Marcos teóricos para o estudo das práticas tecnológicas

Tecnologia é um tema importante para nossa sociedade; tecnologia, e a tecnologia computacional como um caso particular que é vivenciado no programa de pós-graduação em que se desenvolve o presente trabalho, desempenha um papel importante ao criar novas formas de relação entre pessoas e novas formas de viver como pessoa dentro da sociedade. A eloquente, e a meu ver muito otimista, argumentação da relação entre acesso a tecnologia e desenvolvimento humano apresentada no relatório 2001 do Programa das Nações Unidas para o Desenvolvimento (UNDP, 2001) atesta bem esta relevância, e em especial das tecnologias da informação e comunicação e da biotecnologia, temas ali destacados.

Dentro deste panorama, é relevante investigar, de uma maneira crítica, o fenômeno conhecido como tecnologia, como as tecnologias são desenvolvidas e adotadas, e como se estabelece sua relação com as formas de viver em sociedade e seu significado para as pessoas que a encontram. O presente trabalho, ao filiar seu percurso investigativo ao campo dos estudos em Ciência, Tecnologia e Sociedade (CTS), procura compreender de maneira ampla a criação e o uso da tecnologia e do conhecimento científico, e a maneira em que enlaçam inúmeros contextos da sociedade. A inovação tecnológica e a pesquisa científica, nesta perspectiva, são consideradas enquanto forma de agência e reciprocamente como produção social, em uma constituição mútua entre valores sociais, políticos e culturais, e ciência e tecnologia (Hess, 2007; Pfaffenberger, 1988).

Partimos do princípio de que a tecnologia está intimamente entrelaçada à forma como as pessoas vivem. A alimentação, o abrigo, a comunicação, o trabalho, a locomoção e muitos outros fazeres humanos são desempenhados através dos processos e artefatos da

tecnologia. As estruturas escolhidas para instituir e manter estas tecnologias influenciam profundamente todos estes aspectos da vida das pessoas em uma sociedade. A influência disseminada da tecnologia na vida das pessoas, no entanto, não significa que a tecnologia determine estes aspectos da organização social. Não consideramos que a tecnologia é capaz de *especificar* e *determinar* suas formas de uso e os padrões de vida social decorrentes. Uma análise atenta revela que tecnologias são construídas e apropriadas dentro de um contexto social, e que o resultado desta apropriação depende de como as pessoas envolvidas apreendem o campo social reestruturado, como recriam seus objetivos e como mobilizam a partir daí outras pessoas e recursos.

O conceito de tecnologias da inteligência, como apresentado por Pierre Lévy (Lévy, 1993), é útil para esta reflexão. As formas técnicas, aponta o autor, vinculam-se à produção da cultura em relação de mútua constituição. Os artefatos materiais, como suporte ou ferramenta de inscrição, produzem história e preservam em si formas de entender e relacionar-se com o mundo. Citando como exemplo ferramentas, armas, edifícios ou estradas, Lévy argumenta que "a partir do momento em que uma relação é inscrita na matéria resistente... torna-se permanente" (p.76), e assim tanto linguagem como técnica vão constituindo aquilo que uma sociedade sabe de si mesma, sua memória, como percebe a si e a sua história.

Este quadro conceitual é apresentado pelo autor com o objetivo de abordar o fenômeno contemporâneo constituído pela disseminação do uso de uma técnica em especial: a informática computacional e as redes digitais. O interesse central desta abordagem não é definir a priori o que é a tecnologia computacional ou seu potencial. Ao contrário, o autor ressalta o caráter "em formação" desta tecnologia, e as transformações das práticas culturais que vão se vinculando em seu entorno. A programação, a interface e o dado estabelecem-se paulatinamente como objetos e fazeres presentes no dia-a-dia da cultura, mas ao mesmo tempo outras atividades passam a ser mediadas pelo computador: o fazer da imagem fotográfica e videográfica, o fazer da música, o fazer da narrativa escrita. Esta mediação tornou possível novas formas de apropriação destes fazeres, e neste processo vem os transformando profundamente. Alguns pontos desta transformação são considerados especialmente importantes por Lévy, que cita, por exemplo, a mudança nos dispositivos técnicos e substratos para produção de objetos culturais: "a codificação digital relega a um segundo plano o tema do material" (p. 102). Os agentes da produção cultural

também transformam-se, na medida em que habilidades na manipulação computacional tornam-se caminho para entrada no universo da produção, e as redes digitais funcionam como canal acessível para distribuição desta produção. O tempo, no sentido de ritmo em que as os eventos ocorrem, também passou por transformações ligadas ao estabelecimento da produção computacional e das redes digitais; o tempo real, a distribuição e o acesso em um ritmo que os agentes envolvidos percebam-se como pertencendo ao agora da interação uns aos outros.

Uma dimensão provocativa é adicionada à análise da produção de sentido mediada pela tecnologia por Vilém Flusser (FLUSSER, 2002). Refletindo sobre a imagem produzida tecnologicamente, mais precisamente a fotografia, Flusser questiona seu processo de produção e o de seus significados. A imagem, propõe Flusser, não é um objeto final e carregado de significado por uma virtude intrínseca de seu modo de produção técnico. Pelo contrário, sua produção e sua apreciação carregam consigo as marcas do sistema complexo dentro do qual é produzido. Neste sistema, o fotógrafo utiliza um aparelho que é fabricado para funcionar de certa maneira; este fotógrafo também é educado pelo sistema sobre como utilizar este aparelho; e até mesmo a distribuição destas imagens passa por canais que as selecionam e interpretam. Flusser discute a fotografia enquanto forma de entender a relação da sociedade com a produção de cultura mediada pelas tecnologias contemporâneas. Sua filosofia da imagem técnica busca desnaturalizar a produção de objetos culturais mediados pela tecnologia, entendendo-a em conjunto com o sistema dentro do qual ocorre a produção e interpretação. O autor propõe tornar visível o programa que é embutido na tecnologia e nas instituições, isto é, modos de fazer incorporados no substrato material e nas práticas consideradas corretas e naturais para manipulação desta tecnologia – em uma convergência crítica com o conceito de tecnologia da inteligência colocado por Lévy. Neste ponto, seu escrutínio filosófico torna-se uma ética da liberdade para a técnica, colocando-se a questão de por que motivos e por que meios libertar a prática tecnológica de seus programas embutidos. “Liberdade é jogar contra o aparelho” (p.75), desafia Flusser, e o caminho que aponta não é dar as costas ao “aparelho”, ao “programa”, mas tê-los como problema a resolver. Sua resposta, em outras palavras, propõe a reflexão crítica da prática como possibilidade para criar, livremente, com os dispositivos técnicos.

Assumimos, portanto, uma perspectiva que reconhece o processo pelo qual foi sendo conferido, em nossa sociedade contemporânea, um papel privilegiado à tecnologia computacional como lugar de produção de cultura e de subjetividade. Esta perspectiva, ademais, procura o engajamento com as materialidades produtivas da computação não através de um simples uso finalista, mas dentro de um tensionamento crítico das suas possibilidades inscritas e principalmente de suas normatividades embutidas e naturalizadas.

## *2.1 Ciência, Tecnologia e Sociedade (CTS)*

A área de Ciência, Tecnologia e Sociedade (CTS), ou Estudos de Ciência e Tecnologia (*Science and Technology Studies*, STS) estabeleceu-se a partir dos anos 1980 como uma forma de investigar a produção de ciência e tecnologia, examinando-o enquanto processo empreendido pela sociedade e as dinâmicas de sua relação e co-constituição com o mundo social. Seu programa de investigação busca ir além de abordagens mais tradicionais como a de Merton (Merton, 1973), em que o conhecimento científico é discutido a partir das suas instituições e da produção de conhecimento considerado objetivo e universal e que o considera como afastado das contingências do sujeito que produz este conhecimento.

Estudos sobre ciência e tecnologia vinculados a CTS, durante a década de 1980, privilegiaram como conceito de pesquisa a *construção social do conhecimento*, enfatizando como o conhecimento científico, e seu processo de construção de métodos e validação de declarações, envolve negociações de fatores tanto técnicos como sociais. Enquanto que uma visão mais tradicional deste processo encara a ciência e a tecnologia como essencialmente *racionais* em seu desenvolvimento, diversos trabalhos desta época dedicaram-se a examinar as múltiplas considerações que constituem a trajetória ao longo da qual o conhecimento e as tecnologias são estabelecidas. Ponderações acerca de evidência científica, segundo esta linha de pesquisa, encontram-se entrelaçadas a uma série de contingências locais, como o problema da interpretação da evidência, participação e negociação que ocorre entre os envolvidos no processo – não necessariamente apenas

cientistas, processos locais de decisão, e outros aspectos não-técnicos, que participam do que resulta como conhecimento aceito no campo (Hess, 2007).

Esta corrente mais inicial de CTS recebeu o nome de Sociologia do Conhecimento Científico (*Sociology of Scientific Knowledge*, SSK). As principais premissas que a orientaram foram propostas em no trabalho de Bloor (Bloor, 1991), e podem ser expressas como os princípios da *simetria*, da *imparcialidade* (ou *relativismo)* e da *reflexividade*. O princípio de simetria recomenda que o mesmo tipo de causas seja utilizado para dar conta de explicações verdadeiras e falsas produzidas pela ciência, evitando atribuir causas "da natureza" para explicações (posteriormente aceitas como) corretas e causas "sociais" para explicações incorretas. O princípio de imparcialidade, por sua vez, diz respeito a dar relatos imparciais ao descrever o processo científico, com relação a verdade vs. falsidade, racionalidade vs. irracionalidade ou sucesso/insucesso da produção de conhecimento. Reflexidade significa que o mesmo tipo de abordagem explicativa deveria ser usado, tanto para analisar a ciência como a própria sociologia do conhecimento. A importância da SSK para o estudo da ciência e da tecnologia foi o de iniciar discussões *sobre* os processos pelos quais a ciência atinge seus resultados e seu envolvimento com a sociedade, ao invés de assumir, sem discussões, um seu caráter racional e neutro. Estes princípios não foram adotados da mesma forma por todos os estudiosos em CTS nesta fase, tendo sido importantes, no entanto, como um ponto de referência. Ao longo da década de 1990, no entanto, seus pressupostos metodológicos foram revistos, à luz de intensos debates que colocaram como centro a questão de um construtivismo radical, que, ao enfatizar excessivamente a construção social do conhecimento, não abriam espaço em sua teoria para a agência material do mundo ou para o papel da evidência científica. Embora no auge deste debate houvesse propostas de retorno a um realismo simplificado, do tipo que sugere que a ciência descreve o mundo de maneira transparente, o encaminhamento ocorrido foi o de um engajamento crítico da comunidade em torno destas questões, e o surgimento de uma multiplicidade de abordagens para dar conta das diversas facetas do universo da produção científica e tecnológica (Hess, 2007).

As interrogações colocadas pela CTS, a partir de então, desenvolveram-se no sentido de compreender as variadas formas que a relação entre o técnico, o científico e o social pode assumir. Nesta perspectiva, tecnologia não tem simplesmente um "impacto" sobre a sociedade, significando que o desenvolvimento tecnológico ocorre *como* processo social,

sendo, desta forma, uma das forças constituintes do mundo social. Pfaffenberger (PFAFFENBERGER, 1988), entre outros autores, propõe que tecnologia não pode ser considerada de maneira independente do seu contexto de criação; não é frutífero compreender a sociedade como isolada do desenvolvimento das tecnologias e do conhecimento. Deve-se ter em consideração a maneira como as formas do fazer, as ferramentas criadas para este fim e o conjunto de noções sobre o mundo, implicadas neste fazer e nestas ferramentas, *estruturam* a vida das pessoas e como produzem seu dia a dia e sua cultura. Por outro lado, o autor também propõe cuidado com uma maneira simétrica de minimizar a relação entre sociedade e tecnologia, que seria o determinismo tecnológico. Para o determinismo tecnológico, ações dentro do mundo social nas quais a tecnologia intervém são determinadas por esta tecnologia; a história desta tecnologia obedece a critérios racionais e operacionais, e os sujeitos sociais que entram em relação com esta tecnologia têm seu comportamento determinado pelas exigências técnicas. Pfaffenberger argumenta, em contrapartida, que a aplicação da tecnologia e a avaliação de seus resultados não deve ser considerada auto-evidente; a implantação, formatação e difusão de uma tecnologia respondem sempre a uma série de imperativos diversos, em que negociam arduamente, entre outros, mercado, valores sociais e as capacidades materiais que a tecnologia tem de transformação do mundo.

O pesquisador em CTS procura compreender a constituição das práticas tecnológicas e do saber científico tanto ao examinar contingências e flexibilidade interpretativa como também ao descrever como fatos tornam-se estabilizados (THOMPSON, 2005, p. 32). Em sua trajetória investigativa, busca mostrar como a *construção* do conhecimento científico implica um intenso trabalho que envolve instituições, ferramentas e convenções para sustentá-lo. Análises em CTS, desta forma, trazem para a discussão estas múltiplas entidades, mostrando como participam na eficácia do agir científico e tecnológico.

Compreender a tecnologia desta forma significa desafiar uma certa "perspectiva padrão" sobre o tema. Esta perspectiva, como "narrativa mestra" da cultura moderna, é sintetizada e analisada por Pfaffenberger (1992). As premissas desta narrativa fazem parte da visão de mundo dos agentes do campo tecnológico – isto é, engenheiros, programadores, cientistas e técnicos - e constituem o pano de fundo sobre o qual estes agentes constroem seu conhecimento e atividade, não sendo normalmente explicitadas, e sim consideradas a priori, não verbalizadas. Esta perspectiva, segundo o autor, vê a

tecnologia como utilitária e produto direto da necessidade humana, sendo funcional sua essencialidade. Em contraposição a essa essência funcional, que é a razão de ser do artefato ou processo tecnológico e que responde aos desafios encontrados pelo homem em sua relação com o mundo, adornos ou estilos particulares podem modificar a aparência do objeto, como expressão cultural específica de um grupo de pessoas, mas sem alterar sua função primordial. Pfaffenberger, a partir da posição dos CTS, questiona este ponto de vista sugerindo que um exame crítico é capaz de colocar em dúvida a ligação clara entre necessidades "evidentes" e respostas tecnológicas "ideais".

Um exemplo adequado é o caso da roda (Basalla, 1988). Segundo Basalla, a roda foi primeiramente utilizada para fins rituais no Oriente Médio, depois militarmente e só depois para transporte; nas civilizações mesoamericanas pré-colombianas, em contraste, não chegou a ser empregada por causa do terreno acidentado e da falta de animais adequados para tração. Mesmo no Oriente Médio, onde foi inventada, a roda foi abandonada e substituída, posteriormente, pelo camelo como meio de transporte. Para Basalla, uma familiaridade com a ideia da roda "levou os estudiosos ocidentais a subestimar a utilidade de animais de transporte e a superestimar a contribuição dada por veículos com rodas na era que precedeu a substituição da roda pelo camelo". A necessidade é definida através da cultura, e não à sua revelia, ou em outras palavras, a existência da necessidade não deve ser solipsisticamente negada, mas devemos considerar que o ser humano desenvolveu um largo horizonte de objetos e técnicas para cada finalidade objetiva a atingir.

Um outro caso de interesse, em particular para a discussão realizada nesta tese, é o da utilização de computação, que teve efeitos profundos na estrutura e qualidade do que é realizado em nossa sociedade, incluindo trabalho e produção econômica; as transformações são tão notáveis que nos referimos a elas como uma "revolução da informação". Em função destas transformações, é freqüente uma visão determinista, do tipo "impacto da tecnologia", para a qual a utilização de computadores leva, automaticamente, a transformações sociais ou a maior produtividade econômica. P. Edwards (P. Edwards, 2001) mostra, através de um conjunto de casos relacionados à utilização de computadores para desenvolvimento de tecnologia militar e à informatização bancária, como a tecnologia de computação e informação são produtos sociais, que incorporam relações de poder e objetivos e estruturas sociais. Com respeito à tecnologia

militar, Edwards trabalha a partir de dois projetos militares americanos, no pós-guerra, com a meta (não realizada) de produzir uma defesa anti-mísseis nucleares. Embora seu potencial militar fosse mínimo, ambos os projetos produziram um sentido de atividade defensiva, e de mobilização ideológica, que engajou líderes políticos civis, tecnocratas militares e engenheiros em um conjunto de pesquisas que financiou e desenvolveu sucessivos modelos de computadores e tecnologias associadas que entraram em uso difundido posteriormente (entre as quais, por exemplo, memória de núcleo magnético e multiprocessamento). Adicionalmente, a interação entre grupos de pesquisa, entidades responsáveis por financiamento e destinatários finais de pesquisa (os militares) levou à constituição de uma complexa relação interinstitucional, em que os comitês de financiamento foram educados na avaliação de propostas técnicas relacionadas aos desenvolvimentos computacionais então incipientes, e a expandir a expectativa em torno das possibilidades de soluções técnicas, computacionais, para problemas ainda não resolvidos. Por fim, estes projetos, mesmo não tendo sucedido em seus objetivos iniciais, legaram um padrão de comando e controle computadorizado para os sistemas militares americanos.

De maneira semelhante, argumenta P. Edwards, a informatização bancária não ocorreu de maneira linear. Em um caso revisado em seu artigo, o autor descreve o processo de informatização em um banco brasileiro na década de 1980, em que a introdução do processamento computadorizado apenas automatizou algumas rotinas burocráticas, permanecendo em um plano secundário dentro da organização da instituição, em função da resistência de gerentes sêniores. O que ocorreu foi que, para evitar esta resistência, a informatização foi introduzida para automatizar processos repetitivos e pouco visíveis no fluxo de informação do banco; e desta maneira, a automatização passou a ser identificada como uma "função de controle" dos funcionários, e associada aos processos mais aborrecidos da instituição. Um processo distinto, no entanto, caracterizou a informatização de um outro banco (britânico), no qual a computadorização foi introduzida com vistas a reestruturar o trabalho e as carreiras dos funcionários. Antes da informatização, funcionários passavam vários anos dentro do banco, em diversas posições e diversas agências, lentamente aprendendo as competências e rotinas características de cada setor de trabalho. O plano de informatização foi introduzida pela direção da instituição para transformar o aprendizado lento e flexível em

direção a um modelo racionalizado, próximo da produção industrial, em que funcionários aprenderiam rapidamente tarefas rotineiras. A consequência desejada pela gerência era que, a partir desta transformação, a carreira dentro do banco passasse a ser caracterizada por várias faixas; funcionários subalternos perderam a possibilidade de migrar de função e chegar a gerências, as quais passaram a ser prerrogativa de executivos especializados.

Estes casos colocam a forma interativa em que tecnologias – das quais a informatização é um caso – afetam a sociedade através do processos de sua construção social. Tecnologias não são apenas "inseridas" na sociedade, já que trazem consigo modelos específicos de como o contexto social deve funcionar ou organizar-se (P. Edwards, 2001). Por este motivo é importante examinar a relação entre tecnologias e sociedade através de ferramentas teóricas e analíticas que dêem conta desta co-construção, sem resumir o papel da tecnologia à sua utilidade pragmática nem atribuí-la a causa de transformações sociais. Mesmo a distinção entre o que é tecnológico e o que é social não deve ser considerada como dada, por ser culturalmente construída, e por este motivo deve ser analisada em cada contexto (Grint & Woolgar, 1992; Guimarães Jr., 2005).

### 2.1.1 Laboratório: lugar da construção do conhecimento e da tecnologia

Estudos sobre a produção de ciência e tecnologia em laboratórios desempenharam um papel importante no desenvolvimento das bases teóricas e metodológicas dos CTS. A importância da escolha do laboratório como unidade de estudo é destacada, por Knorr Cetina (Knorr Cetina, 2001), por permitir um deslocamento do foco em experimentos, com os procedimentos padrão de teste de teoria prévia a partir do isolamento de fatores variáveis e resultados replicáveis "por qualquer um". Para analisar a ciência e o desenvolvimento tecnológico enquanto processos sociais e com efeitos sociais, sem perder de vista a interação entre o mundo e as competências técnicas dos pesquisadores, o laboratório mostra-se como um lugar de observação mais adequado. Adicionalmente, este é um sítio cujas fronteiras são demarcadas pelos próprios participantes. Segundo a autora, o estudo em laboratórios tornou visíveis a ampla gama de atividades que implica a produção de conhecimento, mostrando como os objetos científicos não são apenas construídos tecnicamente, mas também simbolicamente e politicamente. Exemplos deste entrelaçamento entre o técnico e o social são, por exemplo (2001), as formas literárias de

persuasão presentes em artigos científicos, na constituição de alianças e na mobilização de recursos, pelos cientistas, para a realização de seu trabalho, e na maneira como os resultados científicos e tecnológicos intervém no mundo social.

Alguns trabalhos tornaram-se clássicos dos "estudos de laboratório". Entre estes, o de Latour (LATOUR, 2000), que através de um trabalho de campo em um laboratório de biotecnologia, examinou as práticas cotidianas de um laboratório, e o trabalho de articulação que torna possível as atividades técnicas – mobilizando recursos, desde material de laboratório até a atração de pessoas para trabalhar nos projetos. O trabalho de articulação se estende a uma rede discursiva constituída pelos periódicos científicos, congressos e colegas cientistas, dentro da qual algumas afirmações vão reverberar, ser colocadas à prova, e ganhar status de "descoberta científica", e posteriormente estabilizadas como "conhecimento científico".

Em um trabalho que apresenta ressonâncias marcadas com a presente pesquisa, a antropóloga Diane Forsythe (FORSYTHE, 1993, 1999) pesquisou, como cientista social, os próprios praticantes da Inteligência Artificial, em uma perspectiva de análise que se ocupou das formas diversas e mesmo divergentes como a questão da inteligência e do conhecer eram compreendidas e colocadas em prática. As divergentes formas de abordar estas questões foram frutíferas e úteis para os laboratórios, tanto que estes passaram a apropriar os métodos etnográficos de Forsythe para seus próprios projetos, mas por outro lado causaram conflitos justamente em função destas compreensões diferentes. Abordaremos a pesquisa de Forsythe no capítulo 5, em conjunto com outras questões que relacionam a inteligência artificial e o problema do conhecer.

No Brasil, Sá (SÁ, 2006) analisou um caso em que o universo, delimitado pelos próprios participantes, é um "laboratório expandido": uma estação ecológica, onde trabalha uma equipe de primatologistas. Ali, Sá mapeou entre o trabalho subjetivo dos cientistas, na observação de monos-carvoeiros, e a objetivização produzida, necessária para inserir o conhecimento no circuito científico. Seu trabalho observou a socialização de pesquisadores iniciantes nas práticas cotidianas que permitem o cumprimento de suas tarefas científicas, ao mesmo tempo que permite que se relacionem com o ambiente em que a pesquisa é realizada, incluindo a mata, os habitantes da região, e os próprios colegas em convivência na estação. Práticas científicas, de certa forma, passam a ser constituintes

de um *ser* do pesquisador que também contextualiza sua vida social. De particular interesse em seu estudo são as tensões e negociações surgidas de sua posição como pesquisador *de pesquisadores*, posição na qual preocupações de ordem "técnica" – iria a presença de Sá atrapalhar os cientistas em seu trabalho de observação de monos, dentro da mata? – mesclavam-se a outras, em que o escrutínio de sua pesquisa, associado pela líder do grupo de pesquisa a polêmicas ocorridas na década de 1990, figurava como augúrio de exposição ou de crítica ao trabalho realizado. Sá contornou o impasse de maneira criativa, estabelecendo *rapport* ao mostrar-se pesquisador que utilizava práticas semelhantes e paralelas, como a observação participante, àquelas utilizadas pelo grupo.

## *2.2 O que conta como humano?*

Esta tese destaca, como campo onde o estudo é realizado, a Inteligência Artificial (IA), um ramo de engenharia e uma disciplina da tecnologia e da ciência. Seu objetivo, segundo seus praticantes, é o desenvolvimento de artefatos computacionais que realizem tarefas associadas à inteligência ou à racionalidade; algumas de suas linhas investigam ou procuram replicar a forma como o ser humano desempenha essas tarefas ou esses processos de inteligência e racionalidade, enquanto que outras procuram maneiras mais especificamente computacionais (ou seja, não necessariamente similares à humana) para desenvolver seus sistemas. Ao construir sistemas computacionais capazes de realizar tarefas complexas para as quais normalmente se considera que é exigido o desempenho de inteligência humana, constitui-se em um campo de desenvolvimento tecnológico que chama a atenção por vários motivos.

Suas realizações práticas impressionam, trazendo consigo como efeito um encantamento tecnológico. Demonstram como máquinas são capazes de computar e agir sobre o mundo de maneiras que pareciam, até ontem, terreno restrito da distintiva capacidade humana – realizando atividades que em muitos casos apresentam certo grau de dificuldade mesmo para um humano. As questões que essa disciplina levanta também despertam amplo interesse. Se, por um lado, a IA procura identificar e realizar no

computador o que parece humano, por outro, a própria realização na *máquina* já interroga se o que foi produzido é de fato distintivamente *humano*.

Para entender os conceitos e as práticas das pessoas que procuram replicar o humano através de artefatos computacionais, é imprescindível abordar o problema do que conta como humano. Revisaremos nas próximas seções um pouco da problemática envolvida nesta discussão, através de conceitos mobilizados pelo campo da Inteligência Artificial e Computação Afetiva em suas criações materiais e discursivas.

Estabelecer a categoria "humano" através do que nos faz diferentes leva ao exame de critérios de diferença: em quê somos diferentes, e de quem. Estes critérios tornam-se mais visíveis quando estudamos seres liminares, "quase-humanos", quase-pessoas, sujeitos continuamente à operação de classificação que tenta estabilizar seu status. Podemos contar entre este os nossos semelhantes "naturais", tais como os chimpanzés e outros primatas superiores. Também liminares, sob atenta observação em seu processo de tornar-se ou deixar de ser pessoa, são aqueles objetos construídos tecnologicamente para os quais se atribuem características humanas.

Construir a qualificação "humano" enquanto preenchimento de critérios, entretanto, apenas transfere o problema para que comportamentos de fato contam como válidos para estes critérios (Suchman, 2007, p. 228). Linguagem, um marco da condição humana, é um exemplo. Abelhas demonstram uma capacidade de comunicação complexa e elaborada, realizada através de uma série de movimentos significativos desempenhados diante de outras abelhas da colmeia. Este fenômeno é conhecido como a linguagem da "dança" das abelhas. Há um consenso da comunidade científica em qualificar esta dança como linguagem, por causa de seu caráter simbólico e performativo e por possuir um conjunto de regras observáveis. Uma longa controvérsia a respeito do assunto mostra (Crist, 2004), porém, que há um problema em admitir "linguagem" em relação a insetos: a ideia de "seres inferiores" possuindo linguagem, para os cientistas que discordam desta forma de qualificação do fenômeno, solaparia a distintividade humana dentro de uma taxonomia hierárquica da natureza.

Atribuição de linguagem e agência aos quase-humanos símios revela-se da mesma forma sujeito a disputas de interpretação. Edwards (D. Edwards, 1994) retoma diferentes experiências em que chimpanzés foram ensinados a comunicar-se com humanos (na

maior parte das vezes através de linguagem de sinais). Em um complexo jogo de descrições, a habilidade comunicacional destes animais com humanos é narrada ora como sucesso, ora como insuficiente. Edwards destaca como a mesma descrição é usada em um ou outro sentido: imitação como forma de aprendizado ou, de forma oposta, "mera" imitação rejeitada.

Objetos também são presuntivamente imbuídos de humanidade. Criar objetos com atributos humanos faz parte de uma longa tradição de esforços tecnológicos. Riskin (Riskin, 2003) documenta como Vaucanson, um engenhoso mecânico na França do século XVIII, construiu uma série de autômatos que simulavam processos naturais, entre os quais um "pato que defecava" e um autômato com forma humana que tocava flauta. Ao longo do século XIX e XX, um número de outros projetos abordou a mesma questão, constituindo-se como uma linhagem de investigação à qual as recentes pesquisas em inteligência, vida e pessoa artificial podem ser entendidas como pertencendo. Estes objetos procuravam, através da simulação, investigar os fenômenos que constituíam o funcionamento dos seres vivos e também dos seres humanos (Suchman, 2007, p. 229). Ao reproduzir de maneira prática as hipóteses sobre a fisiologia da vida e a possibilidade de reduzi-la a fundamentos maquinais (mecânicos, químicos ou informacionais), investigavam ao mesmo tempo a fronteira entre a vida e a máquina (Riskin, 2003).

A inteligência artificial é abordada de maneira mais específica por Collins (H. M. Collins, 1990), que procura entender como máquinas computacionais são vistas, pelas pessoas, enquanto dispositivos inteligentes. Para Collins, enquanto que o sistema computacional é construído para executar tarefas abstraídas de sua peculiaridade local, o usuário é quem ativamente enraíza a execução destas tarefas na realidade contingente, realizando ações e atribuindo significados necessários para levá-las a cabo dentro das expectativas dos humanos envolvidos. Em outras palavras, segundo o autor, o que torna estas máquinas funcionais é a inteligibilidade situada, local, e quem a provê são as pessoas. Tais máquinas são concebidas de tal forma que funcionam dentro do contexto social que as rodeia, e as pessoas que as utilizam o fazem de forma a ativamente tornar significativa o resultado de suas ações.

### 2.2.1 Informação

Os autômatos de Vaucanson procuravam simular o ser vivo reproduzindo suas funções características através de mecanismos. Enquanto que a proposta de Vaucanson refletia um paradigma mecanicista do problema do viver, replicantes contemporâneos enfatizam outra maneira de compreender o funcionamento do vivo e do humano: a informação.

Informação no entanto não é uma forma "natural" de compreender o mundo. Figurar a experiência vivida em termos informacionais relaciona-se com um contexto cultural específico de práticas e tecnologias, um contexto gerado a partir do desenvolvimento de máquinas computacionais capazes de agir sobre o mundo através da representação deste como informação. Neste sentido, informação é uma categoria cultural (Dourish et al., 2005), não uma propriedade natural do mundo nem a única forma de conhecê-lo: o espaço e a experiência vividos são construídos como significativos de muitas maneiras diferentes, em diferentes contextos.

Levando adiante o argumento, (Dourish et al., 2005) propõem que considerar a informação como unidade de intercâmbio fundamental da realidade é parte de uma transição conceitual ocorrida no século XX, em que modelos informacionais mais abrangentes, processados por computadores digitais cada vez mais poderosos, foram sucessivamente abrindo as portas para a intervenção computacional sobre o real. Esta transição pode ser observada na inversão da metáfora que explica a relação entre mente e o computador. Os pioneiros da computação, nas décadas de 40 e 50, referem-se ao "computador eletrônico" como um equivalente artificial de um ser humano: *computers* (ou *computors*) eram os funcionários, geralmente mulheres, encarregadas de realizar cálculos repetitivos através de métodos repetitivos pré-estabelecidos (Copeland, 2004). Esta denominação foi utilizada até o fim da década de 1940. O computador artificial foi concebido como uma máquina que reproduziria, dentro de certas limitações, os processos da cognição humana. Nos anos 60 e 70 a metáfora foi invertida: o cérebro passou a ser visto como um computador, e sua atividade consistia em processar informação, baseado em entrada recebida dos inputs-sentidos.

O cognitivismo leva em consideração *operações mentais*, que ocorrem mediando estímulos do ambiente e a conduta do indivíduo como resposta (Suchman, 2007). São estas

operações mentais que, em um processo que as formula como universais e independentes de um sujeito – isto é, abstraídas – são vistas como realizadas em máquinas computacionais. A cognição não é apenas semelhante à computação, nesta forma de encarar o tema; para a abordagem cognitivista, cognição *é* computação, e a inteligência é um fenômeno mental localizado no cérebro e individual.

Esta abordagem possui um contraponto, porém. Garfinkel (1984) argumenta, a partir de uma série de estudos empíricos, que, antes de formas lógicas puras, cognição e racionalidade constituem-se enquanto propriedades testemunháveis da interação social. A interpretação de eventos e atitudes enquanto racionais é construída dentro de interação social. O significado de racionalidade emerge de práticas compartilhadas, ganhando relevância e utilidade na medida em que é demonstrada e utilizada em interação situada. Em outras palavras, ao referir-se a cognição e racionalidade como propriedades "do cérebro", reinscrevem-se características da vida social no modelo ao invés de descobrir características próprias do fenômeno (Dourish et al., 2005).

Analisando o processo de decisão em júris, Garfinkel (Garfinkel, 1984) demonstra como os critérios para decidir sobre o caráter de relatos e evidências e como estes são construídos como fatos ou como aparência são fruto contínuo de negociação, explícita ou implícita, sobre a racionalidade e a adequação dos processos individuais de consideração e pensamento. Os critérios gerais a serem seguidos pelos jurados são socialmente aprendidos no decorrer de suas carreiras como cidadãos e como jurados; mas o ponto fundamental é que a aplicação destes critérios não é automática, pré-significada, mas sim continuamente pesada e avaliada dentro da fundamentação do momento situado, vivido e da avaliação e consideração da questão entre os pares. A racionalidade e a plausibilidade de fatos é constituída em processo, construída cuidadosamente de relatos, evidências e recursos ao senso comum, ao que faz sentido para estas pessoas. Isto torna o procedimento de dar conta da do que aconteceu de fato, isto é, a missão do jurado, não frágil mas pelo contrário mais firmemente enraizado na realidade conforme percebida por um membro competente da sociedade, ou seja, o que seria uma pessoa inteligente e racional.

### 2.2.2 Redes e Híbridos

Esta tese dialoga de várias maneiras com a Teoria Ator-Rede (ANT, Actor-Network Theory) (Callon, 1997; Latour, 1994). Uma das formas de compreender a constituição mútua do fazer tecnológico e da vida social, a proposta da Teoria Ator-Rede centra sua abordagem na forma como nossa vida social é constituída em redes de relações entre humanos e não-humanos. A mesma pessoa participa de diferentes redes, cada qual com tipos particulares de relação entre seus participantes, e constituindo uma esfera distinta da vida social. As pessoas são pensadas como também estabelecendo relações com os artefatos tecnológicos – respondendo a expectativas, normas e funcionalidades que são entendidas como pertencendo ao uso do artefato. A relação com objetos e artefatos cria mediações entre pessoas, estabelecendo-se como parte das relações próprias das redes sociais; e à medida que mais e mais artefatos são utilizados pelas pessoas, e mais e mais relações são estabelecidas com ou através destes artefatos, tanto mais as redes sociais e seu funcionamento passam a ser inteligíveis apenas a partir do ponto de vista que considera a rede integralmente, humanos mais artefatos. A noção central desta perspectiva é a e que a relação não ocorre entre pessoas, mas entre seres compostos de pessoas e os artefatos que as cercam.

Em outras palavras, a teoria ator-rede procura pensar a agência tecnológica reconhecendo que objetos tecnológicos possuem um papel dentro das redes de trocas sociais, o que os constitui como atores sociais. Latour (Latour, 2001) descreve a sociedade tecnológica através da forma como estes entes, chamados por ele de “atuantes”, são cooptados e estabilizados em redes. Latour utiliza o exemplo do micróbio do antraz descoberto por Pasteur, na França do século XIX. No início de sua pesquisa, Pasteur lidava com um agente imprevisível e incontrolável, que causava danos ao rebanho de ovelhas francês, e sobre o qual pouco se podia dizer. No decorrer do trabalho, Pasteur e os franceses puderam passar a interagir com um agente que já tinha nome, e sobre o qual já era possível dizer algo. Esta possibilidade de dizer e de agir estava ancorada na mobilização de uma grande rede de atuantes: laboratórios cheios de equipamentos e laboratoristas treinados, o Estado francês que deu suporte a uma política de vacinação do rebanho, universidades e escolas técnicas nas quais cientistas e laboratoristas eram

introduzidos às práticas que os tornavam capazes de reproduzir e confirmar os resultados de Pasteur.

O nome que Latour dá às entidades assim constituídas é "híbridos". Não se trata apenas de agência humana mediada tecnologicamente. A agência primária humana é alterada e estendida de forma substancial à medida em que estes híbridos são estabelecidos. Ao mesmo tempo, o caráter duplamente contingente das relações com estes atores-objetos é incorporado aos híbridos. Ao entrar em relação com o híbrido, entra-se em relação complexa com estas agências – a ação humana dentro do híbrido já não é "causativa", determinante, mas contingente por causa da complexidade agencial dos objetos não-humanos às quais se associou.

A teoria ator-rede não atribui aos atores-objetos, não-humanos, a categoria de "pessoa". Os atores não-humanos são pensados como agentes porque o conceito de agência perde sua referência antropomórfica, e passa a ser vinculado à contingência mútua, à imprevisibilidade da interação entre o agente humano e o não-humano.

O problema da agência híbrida encontra eco em questões jurídicas, que procuram resolver como atribuir a responsabilidade em atos de contratação entre sistemas informáticos (Teubner, 2006). A discussão gira em torno da concepção do contratante – em que medida é reconhecido o caráter híbrido da agência que efetua o contrato – e reflete-se na prática na forma como é aporcionada a responsabilidade e a possibilidade de rescisão do contrato (por exemplo, em caso de falha de algum dos agentes contratantes).

Ou seja, o híbrido é mais do que a agência humana – mesmo que o humano seja "responsável" perante a sociedade juridicamente falando, o híbrido contém agências não antecipáveis totalmente pelos humanos. O conjunto humano e não-humanos reage de forma diferente do que só o humano; e quanto mais complexo for o arranjo híbrido, tanto menos a agência humana é a única a determinar o comportamento do conjunto.

É nesse sentido que as redes estudadas pela teoria ator-rede não são como redes conectando entidades dadas, simplesmente existentes, mas como a "rede que configura ontologias. Os agentes [e suas características] ... dependem da morfologia das relações em que estão envolvidas (Callon, 1997). As redes que a ANT se propõe a estudar são recurso metodológico, como uma forma de entender a constituição da sociedade, e não

ontológicas (existem redes reais tais como a rede telefônica, mas estas não são o objeto primário desta teoria).

### 2.2.3 Inteligência Artificial: desafio analítico

A Inteligência Artificial constitui-se em um campo tecnológico que é particular desafio para uma análise nos termos propostos pelos estudos de Ciência, Tecnologia e Sociedade. Seus praticantes excursionam com desenvoltura em várias áreas de pesquisa sobre o humano e o social, tais como o conhecimento, o raciocínio, a linguagem, a emoção e as relações sociais. Constroem máquinas que demonstram como possuindo toda esta série de características humanas, e, adicionalmente, reivindicam estas máquinas como ponto de partida e recurso adequado para, por sua vez, estudar estas características. Muitas das análises sobre IA e suas relações com a sociedade terminam, desta forma, por apontar para transformações sociais *causadas* por esta tecnologia, ou, por outro lado, concentram-se em elencar características que a sociedade ou o indivíduo possuiriam e que deveriam ser implementadas pelas técnicas da IA (como é o caso de vários dos projetos investigados no trabalho de campo).

A questão que pode ser levantada é que ambas as alternativas ressentem-se de um certo determinismo tecnológico, conforme discutido acima. Ao abordarem a relação entre a IA e a sociedade, seus proponentes dão primazia à técnica ou aos dispositivos, e elaboram seus argumentos como se estes, assim delimitados, causassem transformações sociais e alterassem, por si, as formas como pessoas se relacionam entre si, com seu trabalho, e com sua cultura. O quadro de referência para estas análises é o do "impacto", em que uma via de mão única leva do objeto tecnológico a efeitos sociais, em um sentido de movimento do interno (de dentro do mundo tecnológico) para o externo (para o mundo social mais amplo).

A segunda alternativa percorre um sentido de movimento de certa forma inverso. A partir de características imputadas à sociedade e às pessoas que a habitam, um equivalente tecnológico é proposto com o objetivo de atuar de maneira semelhante ao humano. O determinismo aqui está na ideia de tomar como entidades estáveis traços do humano e do social cuja coesão e inteligibilidade são fruto de trabalho interpretativo, conceitual. Em outras palavras, participantes da IA tomam conceitos – desenvolvidos

como constructos epistemológicos – e trabalham sobre eles, buscando implementá-los como entidades, de estatuto ontologicamente estável. Entre estes conceitos estão, por exemplo, a emoção ou a relação de sociabilidade: a multiplicidade de teorias científicas sobre estes temas atesta, não a inexistência de seu referente, mas justamente a procura por uma teoria que delimite e torne inteligíveis os fenômenos a que se referem. A busca por definições estabilizadas e que se refiram a entes estáveis poderia ser, inclusive, considerado um dos paradigmas generativos da própria IA, já que é o processo de questionamento através do qual são traçados objetivos a serem alcançados pela pesquisa desta área. A revisão conceitual aqui exposta propõe orientar a interpretação do universo tecnológico que se vai investigar. A Inteligência Artificial e a Computação Afetiva, como outras áreas de realização tecnológica, possuem maneiras próprias de entender e elaborar os recursos de que se utilizam para construir suas realizações. As características humanas que procuram compreender e replicar para tornar seus artefatos mais capazes e eficientes não são exceção. Cognição, afeto, comunicação e outros traços pelos quais nos tornamos humanos e através dos quais nos reconhecemos como tais são formulados, pelos praticantes deste campo, através de uma complexa assemblage de conceitos e técnicas. Seu intuito é explicitamente fazer possível sua implementação, com todas as dificuldades que ocorrem e que são próprias do cotidiano da realização de artefatos, e também tornar esta implementação reconhecível como uma instância da característica que a modelou.

Seja pelas restrições impostas pelas mundanas praticalidades da implementação, seja por questões relacionadas à cultura própria do campo tecnológico (Forsythe, 1993), entretanto, as noções de humano mobilizadas pelos praticantes são recorrentemente reminiscentes da própria situação destas pessoas dentro de seu universo profissional: características humanas são figuradas como análogas de conceitos computacionais, como enunciáveis de forma independente de corpos humanos, como discretas e enumeráveis, como mensuráveis. Talvez mais problemática do que a formulação propriamente dita, entretanto, é a forma como esta é tacitamente considerada *a* formulação possível do humano, tomando a eficiência e o sucesso da tecnologia como evidências de que as concepções empregadas são afins da natureza humana, sem que seja necessária uma discussão a este respeito.

O percurso aqui considerado, então, recapitula várias destas características, e apresenta formas de entendê-las a partir de diferentes pontos de vista. Para tratar

justamente esta questão é que este trabalho é proposto, partindo da observação da realização dos artefatos e da inscrição das marcas de sua presença como construções funcionais, como ponto de partida para levantar formas alternativas de entender o humano dentro desta prática. Ao trazer para a discussão de produção do humanos uma série de contextos diferentes e suas formas de compreensão, sem nunca perder de vista o trabalho real que produz a tecnologia, a intenção é enriquecer o fazer tecnológico e abrir possibilidades que estejam baseadas na reflexão e na inclusão da perspectiva de outros humanos.

# 3 Inteligência Artificial: panorama conceitual e histórico

A Inteligência Artificial (IA) será considerada, para os objetivos deste trabalho, a partir de uma perspectiva proposta por seus praticantes: como um campo de tecnologia e de ciência, como uma forma de engenharia (Luger, 2002; Russell & Norvig, 1995). A Inteligência Artificial propõe, como seu fim, construir artefatos computacionais – programas e sistemas controlados por computador – que realizem tarefas para as quais usualmente é considerado que é necessária a inteligência e racionalidade características do humano.

O computador, enquanto artefato tecnológico, está indissoluvelmente ligado à IA da forma como praticada atualmente. Sua origem insere-se dentro de uma linhagem de criações artefatuais destinadas a automatizar cálculos. Computador ("computer" ou "computor") era o funcionário encarregado de realizar cálculos, aplicando rotinas estabelecidas para obter os resultados a partir dos números de entrada. Máquinas de calcular mecânicas, eletromecânicas e analógicas eram utilizadas nas décadas de 20 e 30, por estes funcionários, para obter resultados numéricos mais rapidamente do que os cálculos efetuados à mão (Copeland, 2004, p. 75; P. Edwards, 2001, p. 75). É nesse sentido que deve ser lido o trecho de Turing (1950): "O comportamento do computador em qualquer momento é determinado pelos símbolos que ele está observando, e seu 'estado mental' neste momento"[2]. Os computadores eletrônicos foram desenvolvidos, a partir do fim da década de 30, como máquinas que fossem capazes de efetuar cálculos a partir de rotinas estabelecidas e em um número finito de passos. Eventualmente, o desenho típico

2 "The behaviour of the computer at any moment is determined by the symbols which he is observing, and his 'state of mind' at that moment"

de um computador do tipo como conhecemos, eletrônico, digital e programável, estabeleceu-se com os computadores comercializados pela IBM a partir de 1952, baseados em válvulas.

Uma outra dimensão do computador é contemporânea a esta, e está intimamente ligada ao seu desenvolvimento como artefato utilitário que realiza cálculos. Uma série de desenvolvimentos que interligaram descobertas em matemática, fisiologia e engenharia elétrica, além da novidade que era o computador eletrônico, deram origem a um campo de estudos que se denominou cibernética[3]. A cibernética partia da premissa de que o sistema nervoso de animais era funcionalmente análogo aos circuitos digitais que compunham as novas máquinas de calcular. Sua abordagem era a de que, em ambos os casos, os fenômenos essenciais a serem entendidos eram o 'controle' e a 'comunicação' operados a partir da 'informação', e que o substrato físico em que estes fenômenos ocorriam poderia ser abstraído:

> É digno de nota o fato de que os sistemas nervosos animal e humano, sabidamente capazes de um sistema de computação, contêm elementos [i. e., os neurônios] idealmente adequados para atuar como relés (Wiener, 1970, p. 171)

A cibernética introduzia uma teoria e uma perspectiva de entendimento do mundo em que a informação era a questão central, através da noção de realimentação em sistemas biológicos e eletrônicos não-simbólicos, e construía uma ponte analógica entre a natureza e o artefato, com a intenção de que um pudesse modelar o outro. Fenômenos como a oscilação eram equiparados e explicados simultaneamente nos mundos da mecânica, da eletrotécnica e da fisiologia nervosa, utilizando um único modelo matemático. Dentro deste contexto, o cérebro e o computador eram modelos um do outro: 'uma grande máquina de calcular... na forma de aparelho mecânico ou elétrico ou na do próprio cérebro' (Wiener, 1970).

Nesta mesma época, Alan Turing discutia, em um artigo que veio a ser muito influente (Turing, 1950), uma ideia de inteligência de máquina. Turing propôs que, se uma máquina fosse capaz de realizar determinado comportamento de tal forma que para um observador ela passasse como humana, então esta máquina poderia ser dita inteligente.

---

3 Para um testemunho histórico deste momento (1947), ver (Wiener, 1970), prefácio à edição original.

Formulada como um "teste", a proposta incluía um tipo de ação considerada pelo autor como critério adequado para "inteligência", e veio a ser conhecida como o "Teste de Turing". O comportamento que Turing sugere para exercitar esta determinação é um diálogo de perguntas e respostas, realizado através de uma interface escrita datilografada (hoje, digitada), para que sinais físicos como aparência e voz não influenciassem o julgador. As máquinas seriam computadores digitais, "capazes de realizar qualquer operação que poderia ser feita por um computador humano" (p. 436). O argumento de Turing era que, se um ser humano, interrogando um outro humano e uma máquina, não fosse capaz de distinguir qual é a máquina, então a máquina deveria poder ser descrita como inteligente. Mais exatamente, Turing afirma que a resposta ao desafio proposto poderia substituir a questão "Podem máquinas pensar?". Colocando-se dentro de uma tradição dualista, cartesiana, que figura a cognição como fenômeno isolável, e que seria muito influente dentro do que se constituiria como a Inteligência Artificial, Turing afirma que o teste, da maneira como proposto, tinha a vantagem de "traçar uma linha nítida entre as capacidades física e intelectual de um homem" (p. 434), sendo adequado para tratar "quase qualquer campo da empresa humana que desejaríamos incluir" (p. 435).

A Inteligência Artificial originou-se neste contexto. O termo surgiu em 1956, quando foi promovido um seminário em Dartmouth (Estados Unidos), propondo um período de estudos de "inteligência artificial", "com base na conjectura de que qualquer aspecto de aprendizado ou outra característica de inteligência pode em princípio ser tão precisamente descrita que uma máquina pode ser feita para simulá-la" (Russell & Norvig, 2010, p. 17). Deste seminário participaram pesquisadores que viriam a ser eminências dentro do campo que se estabelecia, como John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon. Estes dois últimos apresentaram ali um programa de raciocínio lógico – *Logic Theorist* – que manipulava listas de símbolos e resolvia proposições lógicas. Logo depois, apresentaram uma nova versão, o GPS – *General Problem Solver*, "projetado desde o início para imitar protocolos humanos de resolução de problemas", "provavelmente o primeiro programa a incorporar a abordagem 'pensar como humano'" (Russell & Norvig, 1995, pp. 16-17).

O início da Inteligência Artificial, durante as décadas de 1950 e de 1960, concentrou-se em questões de raciocínio lógico-matemático automatizado e de representação simbólica de conhecimento, dedicando-se à criação de sistemas computacionais que

fossem capazes de resolver problemas de raciocínio lógico. Durante este período, foi importante a transição pela qual passou-se a considerar importante o ponto de vista pelo qual a rede de neurônios que compõe o cérebro é um "processador de informação": redes de neurônios como dispositivos que armazenam e transformam informação (em contraposição a um ponto de vista bioquímico-funcional). Durante este mesmo período, é interessante observar que também a genética passou a destacar a informação – codificada no código genético, replicada pelo mecanismo celular – como ponto de vista fundamental (STEELS, 2007).

Durante os anos 1970, as técnicas de representação e conhecimento foram aplicadas à construção de sistemas de IA com a possibilidade de utilização prática. Estamos nos referindo aos Sistemas Especialistas, também chamados de sistemas baseados em conhecimento, que são programas de computador que codificam, em uma *base de conhecimento*, conjuntos de conhecimento, como fatos, relações e regras, de disciplinas específicas (Harmon, 1985; Russell, 1995). Sistemas desse tipo começaram a ser desenvolvidos nesta época, nos Estados Unidos, com aplicação em áreas de especialidade tais como diagnóstico médico (o sistema MYCIN) e prospecção geológica (sistema PROSPECTOR). Seu funcionamento baseia-se na codificação, ou, segundo a linguagem de seus praticantes, na *representação* de *fatos* e de *regras* que descrevem o conhecimento da área em questão, bem como na aplicação de inferência lógica no encadeamento dessas regras, para chegar a uma afirmação que responda ao problema proposto. Tais sistemas, segundo seus proponentes, são capazes de realizar tarefas que requerem conhecimento específico de uma área, em um nível próximo ou superior ao de um especialista (HARMON, 1985).

O conjunto codificado de conhecimentos, que caracteriza e é ponto de partida para a operação do sistema especialista, é chamado de base de conhecimento. A base de conhecimentos é construída por profissionais chamados de "engenheiros de conhecimento", a partir da coleta de informação fornecida por especialistas humanos, ou por fontes documentais. Essa informação é, em seguida, codificada e sistematizada – representada – em uma linguagem lógica ou computacional específica, como regras e procedimentos, e tais regras e procedimentos são representados em linguagens computacionais adequadas. Um sistema completo típico constitui-se da base de conhecimento, de um módulo lógico que é responsável por encadear as regras

adequadamente para chegar a uma resposta, e da interface através da qual a tarefa é codificada para ser apresentada ao programa e a resposta é publicada.

Posteriormente, outras abordagens emergiram, buscando explorar alternativas tanto em relação aos problemas a serem resolvidos, como em relação à introdução de técnicas diferentes. Uma dessas abordagens é o conexionismo, que retomou destaque na década de 1980, e que procura resolver problemas de maneira não-simbólica, através da implementação de entidades computacionais similares a redes de neurônios conectados entre si. A origem do conexionismo (ou redes neurais, como também é chamado) está no modelo matemático do neurônio que McCulloch e Pitts (McCulloch & Pitts, 1943) apresentaram, sendo usada para produzir uma rede neural artificial (em computador) chamada de *perceptron*. O modelo do perceptron é constituído de uma rede de unidades de cálculo, implementada em computador e organizada em duas *camadas*, a camada de entrada e a de saída. Cada unidade da camada de entrada recebe um sinal externo, e está conectada a todas as unidades da camada de saída, para as quais propaga um sinal, cada sinal sendo modificado por um *peso* de 0 a 1. Seguindo o modelo (simplificado) de McCulloch e Pitts, o sinal propagado por uma uma unidade é zero se a soma dos sinais que recebe é maior que um certo valor de ativação. Redes neurais funcionam da seguinte forma: apresenta-se uma entrada, composta por um conjunto de sinais, e verifica-se a saída, composta pelo conjunto de sinais presentes nas unidades da camada respectiva[4]. Se a saída for diferente da esperada, há uma regra para calcular novos pesos para os sinais propagados. Este processo, chamado de *treinamento*, é repetido várias vezes até que a saída seja semelhante à esperada, quando então a rede pode operar para entradas "novas" com resultado presumivelmente correto.

Redes neurais artificiais simples como as descritas são capazes de dar conta de alguns problemas interessantes, mas em função de limitações matemáticas seu desenvolvimento foi abandonado nos anos 1960. No entanto, durante os anos 1980, diversos avanços teóricos permitiram novos desenvolvimentos, incluindo redes neurais em várias camadas e técnicas de treinamento mais potentes (Rumelhart et al., 1986). Poder de processamento maior em computadores também foi um fator que facilitou e ampliou o uso e

4 Há redes neurais (como as de Hopfield) construídas de maneira que todos os nós são entrada e saída (Russell & Norvig, 2010)

desenvolvimento desta abordagem, empregada por exemplo para classificação de padrões em dados, controle difuso, aproximação de funções e restauração de informação faltante (Li, 2008, p. 35). O ponto mais interessante, no entanto, desta abordagem, é que ela pode manipular informação de maneira não simbólica: a obtenção de dados na saída depende da presença de um conjunto de valores em unidades nas camadas intermediárias, que não correspondem a uma representação proposital ou simbólica de algum dado[5], como ao contrário é o caso das abordagens baseadas em lógica e representação (simbólica) do conhecimento.

Ligada ao conexionismo está a robótica incorporada ou situada (Steels & Brooks, 1995), em que a representação simbólica seria substituída por manipulação mais direta dos dados do ambiente. Foi desenvolvida também a partir de meados dos anos 1980, em contraponto à IA mais tradicional que é fundada na representação abstrata, simbólica de conhecimento objetivo e explícito e na manipulação lógica deste conhecimento, em separado da interação complexa com o mundo, os proponentes da robótica situada afirmam que estes princípios não dão conta de muitos aspectos do que constitui inteligência ou comportamentos observáveis em seres vivos. Propõe, em substituição, princípios de *situação*, *corporeidade* e *emergência*. Sistemas *situados* não dispõe de representação interna prévia do mundo modificada durante o transcorrer de seu funcionamento, mas ao contrário o sistema refere-se continuamente aos seus sensores, diretamente – segundo Brooks, "usando o mundo como seu modelo" (Brooks, 1995, p. 54). A referência contínua ao mundo implica que o sistema possui uma presença neste mundo, não sendo portanto um processamento abstrato e simbólico ou virtual; *corporeidade*, por isto, é o motivo pelo qual os sistemas propostos por esta abordagem são robóticos, equipamentos que estão dentro do mundo, em um encontro e não em separação por abstração ou modelo processado independentemente. E, por fim, comportamentos do sistema são esperados como *emergindo* da interação entre módulos simples a partir dos quais é construído, e não pré-programados e planejados de antemão. A premissa por trás dessa proposta é que interação com o ambiente é a origem de comportamentos complexos, em sistemas biológicos usados como comparação e inspiração, e que inteligência, mais do

5 Embora existam sistemas construídos em redes neurais que atribuem correspondência entre estados de nó e símbolos.

que uma capacidade completa de processamento de conhecimento explicitamente representado, está ligada a essa dinâmica.

A robótica situada apropria-se, para construir sua proposta, de motivações teóricas e soluções práticas de redes neurais e da cibernética, cujo desenvolvimento, desde os anos 1940, ocorreu à parte da corrente majoritária da IA. Possuem em comum o recurso ao processamento de informação de uma maneira que não abstraía modelos simbólicos. A cibernética, que passou a ser conhecida por *teoria de controle* (ou *sistemas de controle*), contribuiu com noções de controle de sinais analógicos complexos, incluindo predição de comportamento destes sinais, através da procura de modelos matemáticos lineares e não-lineares, e da realimentação, isto é, a presença de informação de saída como componente da entrada do sistema. A teoria de redes neurais trouxe a ideia de sistemas computacionais que processam a informação não previamente organizada em modelo explícito.

Pesquisadores em IA voltaram-se também para o tratamento de problemas que envolvem dados conhecidos de maneira parcial, probabilística ou incerta. Abordagens anteriores baseadas em lógica encontravam dificuldades quando aplicadas a problemas que envolvem conhecimento não completamente especificado. Especificar e modelar completamente um conjunto de conhecimentos pode não ser factível, por ser o modelo incompleto ou sujeito a ocorrências não previsíveis ou aleatórias fora do alcance da ação do sistema. De fato, este é um reflexo da inexauribilidade do mundo como fonte de traços inteligíveis ou intervenientes. Garantir um resultado "verdadeiro" do ponto de vista lógico, nestas situações, não é possível, embora, se examinados à luz de práticas humanas, um resultado "adequado" seja possível. Uma das soluções propostas foi a da utilização de encadeamentos lógicos *probabilísticos*, para os quais se torna possível representar e calcular, dentro de um rigor lógico, possibilidade (probabilidade) de estados, simbolicamente abstraídos, de um sistema, desde que as probabilidades de alguns estados chave sejam conhecidos. Conhecida como *redes bayesianas* (Luger, 2002, pp. 334-344; Pearl, 1991, 1996), esta abordagem trabalha com a representação de nós em uma rede, em que cada nó corresponde a um estado modelado do problema original. Sistemas de IA utilizando estas técnicas (e outras intimamente relacionadas) são construídos para prover *decisão* em um contexto que pode ser modelado, mas para o qual a ocorrência de mudanças não pode ser exatamente prevista.

Outra linha de trabalho relacionada busca tratar, ao invés de estados discretos já interpretados em um modelo do mundo, conjuntos amplos de dados numéricos ou numerizáveis (como imagens) em que há incertezas de diversos tipos. Tratamentos matemáticos que lidam com incertezas, como conjuntos difusos (lógica *fuzzy*), conjuntos aproximados e modelo de nuvem (Li, 2008) são utilizados para, a partir da relação entre matrizes numéricas de medições, produzir dados mais fundamentais, não evidentes na matriz original, que mesmo assim reflitam as incertezas associadas ao próprio problema. Li (2008) dá exemplos de aplicações para *data mining* (garimpo ou mineração de dados) em dados geográficos, de bolsas de valores, e em reconhecimento de faces.

*Agentes inteligentes*, ou *agentes* artificiais, por fim, são uma estratégia peculiar de pensar e desenvolver sistemas de Inteligência Artificial que são de interesse para o presente trabalho. Agentes inteligentes podem ser definidos como um artefato individual (um software, ou mesmo um equipamento robótico) que possui entradas de dados, através das quais recebe um fluxo de informação proveniente do ambiente, e saídas que efetuam ou atuam de alguma maneira sobre o ambiente (Luger, 2002; Russell & Norvig, 1995). O desenvolvimento de sistemas baseados em agentes constitui-se em uma parcela significativa da pesquisa em IA, incluindo os grupos, brasileiro e europeu, em que foi realizado o trabalho de campo. Há, de maneira geral, duas abordagens para agentes inteligentes: a primeira, que é a utilizada nos grupos pesquisados, descende da linha de IA de raciocínio simbólico, e é aquela sintetizada na obra de Russel & Norvig. Caracteriza-se por abstrair raciocínio lógico simbólico como traço essencial da inteligência humana, como em McCarthy & Hayes (1969), equiparando sistemas computacionais e humanos como instâncias de agentes inteligentes, na medida em que presuntivamente atendam aos requisitos desta abstração. A outra abordagem vincula-se à robótica situada, e sua distintividade está em propor sistemas de IA compostos por subsistemas autônomos, *situados* na acepção usada por Brooks (1995), voltados cada qual para funcionalidades/operatividades específicas. O sistema como um todo seria resultado da interação entre estes subsistemas específicos. Ambas as abordagens enfatizam a reatividade e a autonomia (relativa) de cada agente, significando que, mais do que simplesmente processar lotes de informação, agentes continuamente têm acesso a dados originados em seu ambiente (que pode ser um ambiente puramente informático, no entanto), e continuamente atuam sobre o ambiente em função de decisões calculadas

internamente ao agente. Considerar o humano e o agente inteligente computacional de maneira semelhante e simétrica é uma decisão proposital e de importantes consequências para a forma como a IA entende e constrói artefatos que eventualmente vão se encontrar com pessoas.

A Inteligência Artificial mantém um tenso diálogo com uma série de outras disciplinas, gerando discussões e influenciando (e sendo influenciada por) posturas teóricas que atravessam, entre outras áreas, a filosofia, a biologia e a psicologia. Um exemplo é a ciência cognitiva, que ao investigar o pensar e o conhecer humanos, intercambiou ao longo de sua trajetória, com a IA, com a biologia e com a filosofia, vários conceitos e metáforas explicativas: o pensamento como ato computacional, o ato cognitivo como resultado da interação entre elementos neurais unitários interligados, a consciência como propriedade emergente de um sistema complexo. Para dar conta dos casos aqui apresentados, engajar-nos-emos sistematicamente com contribuições de outras disciplinas que ampliam o horizonte da IA, com o fim de trazer para a cena outras maneiras de compreender e discutir o humano e o que é encenado na máquina como característico do humano.

## *3.1 Computação Afetiva*

Ao longo da década de 1990, a insistência da Ciência da Computação e da Inteligência Artificial no traço cognitivo, racional e lógico como descrição plena do humano deu lugar a um questionamento sobre que outros traços descrevem pessoas, e como estas disciplinas deveriam colocar-se diante destes traços. A Computação Afetiva surgiu neste contexto, como uma proposta de imbuir computadores com competências emocionais: "reconhecê-las, expressá-las e em alguns casos, tê-las", como apresenta Rosalind Picard (Picard, 1997), uma das proponentes originais deste campo. A computação afetiva, conforme formulada por Picard, parte das premissas de que a emoção desempenha uma função essencial para a cognição, e que a emoção possui componentes corporais (fisiológicos) e cognitivos.

A referência neste assunto, para os praticantes da computação afetiva, é Antonio Damásio, em sua obra O Erro de Descartes (DAMÁSIO, 1996). Além da importância da emoção como constitutiva da inteligência, Damásio também destaca a emoção como fenômeno corporal, como um processo encenado pelo corpo e no corpo, como um todo – incluindo o sistema nervoso e o cérebro – e que reciprocamente é percebido também corporalmente, emergindo na cognição a partir do fenômeno corporal. Ligando estas duas premissas, o autor conclui então que não é possível separar corpo e mente como entidades independentes, nem considerar a inteligência, o conhecer como uma faculdade imaterial não situada no mundo físico, propostas tradicionais na filosofia do conhecer estabelecida a partir de Descartes (*cogito ergo sum* [penso logo existo]: o pensar como atividade de uma mente sublimada, imaterial, definindo a existência do sujeito). Esta conclusão que se opõe à cognição "pura" da perspectiva cartesiana é que justifica o título de seu livro, "O Erro de Descartes".

A Computação Afetiva apropriou-se desta abordagem sobre a emoção de uma maneira peculiar e distintiva. Em primeiro lugar, o reconhecimento da emoção como integrante e constitutiva da inteligência e do processo de tomada de decisão pelo sujeito tornou não somente possível, mas relevante o próprio projeto intelectual da Computação Afetiva em uma espécie de contraponto ou complemento necessário à Inteligência Artificial tradicional – ver as considerações de Picard sobre o Teste de Turing e emoção (PICARD, 1997, pp. 12-14). Em outras palavras, para os proponentes da Computação Afetiva, como a Afetividade é essencial para a Cognição, a Computação Afetiva tornou-se rota de passagem para chegar a uma Inteligência Artificial funcional. Mas há dois traços que são importantes para entender como o campo da Computação Afetiva coloca em ação a conceituação de Emoção como constituinte de uma estratégia de construção de tecnologia computacional.

Estes traços são visíveis como relação colocada, pelos praticantes do campo, entre o cognitivo e o corporal na compreensão da emoção. Respondendo em parte ao questionamento colocado por Damásio, e em parte a epistemologias cognitivas do afeto, os praticantes da Computação Afetiva dicotomizam o fenômeno emotivo no binário cognitivo vs. corporal. O primeiro traço que decorre desta situação é a Emoção gerada por uma avaliação de natureza cognitiva do sujeito, isto é, a perspectiva de que emoção ocorre no sujeito como uma consequência da avaliação intelectual que o sujeito faz de uma

situação. Um exemplo de avaliação cognitiva, dentro desta proposta, seria: dada uma certa situação, "não atingir o objetivo" seria uma avaliação que o sujeito faria sobre a situação; e, como consequência da avaliação, a emoção vivida pelo sujeito seria "frustração".

O segundo traço, de certa forma um espelho do primeiro, é a construção do fenômeno emoção como categoria mensurável no corpo da pessoa. Este traço vincula-se à noção de que a emoção pode ser comunicada, sendo percebida através dos sentidos (Picard, 1997, p. 165). "Padrões de informação" são comunicados, cujo lugar de enunciação é o corpo; logo, deve-se buscar no corpo a diferença, a alteração física que significa a emoção. Nesta perspectiva o corpo emotivo é visto através de acontecimentos fisiológicos, acessíveis à mensuração, e procura-se relacionar o acontecimento fisiológico com o emotivo. Picard passa em revista vários trabalhos envolvendo emoção e alterações fisiológicas (1997, pp. 142-164), destacando entre outras sinalizações corporais medidas relacionadas ao batimento cardíaco, à pressão sanguínea e à respiração.

A figuração da emoção como fenômeno mensurável é analisada de maneira crítica por Otniel Dror (Dror, 2001), colocando esta forma de estudar os afetos dentro de uma perspectiva histórica. Sua revisão aborda a longa linhagem dos estudos laboratoriais sobre as emoções que inicia nos fins do século XIX, mostrando como a vinculação desta a uma expressão numérica, mediada por máquinas, confere uma legitimidade científica e permite que o afeto possa entrar em uma variedade de contextos sociais públicos em que, de outra forma, sua demonstração e exposição não eram considerados convenientes. Desta maneira, ao mesmo tempo em que a cultura pós-vitoriana introduzia regras de convívio social que visavam a diferenciação entre espaços de trabalho e espaços privados ou de lazer, junto com um regime mais estrito de expressão ou restrição emocional, ocorria uma transformação conceitual sobre a própria emoção, que tornava-a, pela mediação tecnológica e científica, um fenômeno próprio sobre o qual era possível discurso objetivo. Dror ressalta que esta discursividade sobre o emocional era mais ampla do que unicamente científica, como atestam inúmeros artefatos a circular entre o público desde os anos 1920 e que eram destinados a, de uma forma ou de outra, revelar e medir "estados emocionais": emotógrafos, medidores e detectores "do amor", "de mentiras" e outros. Esta tradição tecno-emoto-gráfica é apresentada por Dror, estendendo-se em uma continuidade até os dias atuais em que, por exemplo, através da internet oferece-se a utilização de algoritmos para calcular as possibilidades de amantes potenciais.

A computação afetiva, quando julga necessário, lança mão de distinções entre o que considera diversos fenômenos afetivos, em função de características como a duração do quadro temporal em que se inscreve o afeto (por exemplo, um estado afetivo de longa duração pode ser denominado *disposição* [mood]), ou como é pensado o locus de encenação deste (por exemplo, *sensações* como ligadas a estímulos sensórios) (PICARD, 1997). O foco, nesta tese, será sobre as emoções, através do percurso investigativo e das propostas apresentadas principalmente por Ortony et al. (1990) e Picard (1997), na medida em que aparecem com destaque como referencial para o trabalho dos grupos com os quais foi realizado o trabalho de campo.

## *3.2 Interação*

A interação com a máquina é locus adequado para examinar concepções do universo tecnológico sobre o humano. Interação, originalmente descrevendo um ato entre pessoas, supõe troca comunicacional e inteligibilidade mútua, uma compreensão compartilhada sobre o ato e sobre o objeto da troca de ações. Ao utilizar esta mesma palavra para descrever a experiência humana frente a sistemas computacionais, é preciso questionar como dar conta desta expectativa de inteligibilidade mútua, agora instanciada entre a pessoa e a máquina.

Como problema prévio a este, podemos colocar a própria questão da interação entre pessoas (SUCHMAN, 2007, p. 34): de que maneira seres humanos conduzem seus processos interacionais, e de que maneira dão conta da expectativa de compreensibilidade entre as partes. Segundo Suchman, há três razões que tornam os artefatos computacionais plausivelmente "interativos". A primeira razão seria sua reatividade: ao humano envolvido no processo são dados meios de controle, incluindo a capacidade de interromper e modificar a resposta da máquina. A conduta da máquina não é aleatória, e sim responde a um projeto, a uma intenção do projetista, o que confere a esta conduta um caráter de pertinente a um propósito.

Os meios de controle mencionados têm ganhado contornos linguísticos, no sentido de que operar estes artefatos consiste em selecionar ou compor ações enunciadas, seja por

escrito ou até mesmo sonoras, como articulação de uma forma de uso da linguagem que permite acesso a operações da máquina. Esta é outra razão pela qual Suchman vê a ideia de uma interatividade computacional ser aceita. Especificando expressamente como linguagem as formas de operar da máquina para o humano que a opera, os projetistas utilizam termos emprestados da linguística para apontar o que acontece entre a máquina e o humano. Isto é, palavras tais como "diálogo" e "subentender" passam a ser utilizadas descrever este processo, trazendo junto consigo um conjunto implícito de suposições sobre propriedades que usualmente eram esperadas de fenômenos comunicacionais (entre humanos).

Além da reatividade e dos meios linguísticos, Suchman aponta para a relativa opacidade dos dispositivos computacionais como chave para seu entendimento como interativos. Opacidade significa não uma ininteligibilidade por insuficiência técnica do participante humano, mas a irredutibilidade da conduta da máquina como resposta a um evento único, localizado. A conduta da máquina computacional é complexa, e sua reação local responde a um conjunto de eventos que constituem seu histórico de operação, e que não se esgota em especificação prévia. O participante assume um design racional para o sistema, e baseia-se em um senso comum sobre o sistema e suas adjacências de significado para antecipar e entender a conduta da máquina. O fato desta conduta não ser completamente especificável, mas ser predizível e compreensível até certo ponto, soma-se às outras razões para construir a percepção de que o sistema computacional não é apenas uma ferramente, mas sim um artefato mais complexo com o qual é possível entrar em interação.

## *3.3 Inteligência Artificial e CTS: para uma análise crítica*

O objetivo de examinar o processo de co-constituição entre tecnologia e sociedade e em rever, a partir de uma postura interpretativa, conceitos e noções normalmente tidos como estabelecidos tanto a respeito de uma como de outra faz de Ciência, Tecnologia e Sociedade (CTS) um campo de estudos especialmente interessado em abordar aquelas áreas tecnológicas que reivindicam para si o desenvolvimento de artefatos *semelhantes ao*

*humano*. Estas áreas tecnológicas, notadamente a Inteligência Artificial, utilizam, para descrever os artefatos que desenvolvem, atributos tais como "raciocínio", "conhecimento", "inteligência", "comunicação", que usualmente eram restritos *e distintivos* do humano.

O humano visto por nossa sociedade é o centro e o ápice das formas de existir. Quer seja o relato de fundo religioso, quer seja a descrição científica, vemo-nos como a mais importante, evoluída e complexa forma de vida existente. Dada a obsessão tecnológica desta mesma sociedade, que investe incessantes esforços para realizar artefatos que sejam funcionalmente cada vez mais admiráveis, causa pouca surpresa o fato de que replicar tecnologicamente capacidades humanas esteja presente como objetivo e marco de sucesso. Engenheiros, cientistas da computação e programadores, agentes do campo que produz a tecnologia da IA, investem o esforço de suas vidas para criar artefatos computacionais dos quais possam dizer que possuem "características humanas". Sistemas especialistas, tutores inteligentes, bots interativos e robôs humanoides, entre outros, são artefatos que imbuem premissas e concepções dos agentes do campo tecnológico sobre o que é um "humano". Como quadros pintados, retratos, reconhecemos a semelhança destes artefatos conosco. Ao mesmo tempo, se observados atentamente e com um certo distanciamento, estes retratos também revelam que este artefato humanizado é modelado a partir de uma visão particular de ser humano, a visão dos agentes do campo tecnológico.

# 4 Questão de pesquisa e desenvolvimento metodológico

A Inteligência Artificial coloca-se um tanto à parte, dentro das ciências com caráter técnico, em função de seus objetivos e de suas premissas de trabalho. Não pretende "apenas" construir ferramentas úteis, adequadas e potentes; pelo contrário, toma um modelo de *ser* humano bem específico, o de pensar racionalmente, e investiga este modelo ao mesmo tempo em que constrói artefatos que jogam com este modelo.

De maneira semelhante a outros temas relacionados ao humano, aqueles abordados e desenvolvidos pela IA não estão pacificamente estabelecidos; inteligência, por exemplo, é um conceito continuamente reinterpretado e recolocado em questão. Isto não causa dificuldade para os praticantes desta ciência técnica; pelo contrário, eles mesmos participam destas reinterpretações, por vezes propondo novos artefatos, por vezes propondo novos modelos que desafiam noções existentes.

A IA é um campo muito atraente nos contornos de suas proposições e de seus desafios: encontrar maneiras sistemáticas de construir artefatos capazes de realizar aquilo que, usualmente, atribuímos como específico da inteligência humana (Russell & Norvig, 1995). O atributo "inteligente", embora atribuído com certa prodigalidade a muitos sistemas controlados por computador, em geral significa um toque a mais de funcionalidade, algo que ultrapassaria o simples "processamento de dados" utilitário, uma certa atenção a contextos diferentes. Não apenas produzir dados, mas interpretá-los ou marcá-los segundo categorias que são fruto de nossa interpretação ou conhecimento do mundo, esta pode ser vista como a distintividade buscada pelos praticantes da IA.

Tendo nascido próxima à Ciência da Computação, a IA permanece ligada à história dos computadores, na medida em que os seus sistemas são realizados em computadores. Em alguns destes casos, um processamento sofisticado dos dados é o objetivo, com a coleta e a categorização através de procedimentos matemáticos que possibilitam a produção de informação não prontamente visível: garimpar, minerar massas de dados, e processá-los por algoritmos que conseguem lidar com dados não perfeitamente exatos ou perfeitamente completos, e por fim separar e apresentar classificações que surgem como resultado.

As suas conjecturas e investigações, no entanto, correm em várias direções – como gotas de mercúrio derramadas – interrogando facetas do humano disponíveis para a curiosidade dos seus pesquisadores. Nestas várias interrogações, a IA por vezes encontra e co-desenvolve-se com outras ciências recentes, com as quais troca metáforas e visões de mundo, das quais talvez o melhor exemplo seja o das ciências cognitivas. O computador, inicialmente um análogo do cérebro, passou a ser o modelo generativo para compreender o pensamento, que, então, tornou-se processamento dos dados fornecidos pelos sentidos: do cérebro eletrônico, passamos ao computador biológico.

Em outras situações, as ciências que estudam o humano são procuradas para ampliar o escopo de funcionalidades dos artefatos desenvolvidos. A inteligência racional, os raciocínios, seriam apenas pontos de partida, e outras formas de *ser* o humano são consideradas, em uma busca para completar os sistemas, de torná-los mais funcionais, mais úteis, mais poderosos. Humanos têm emoções, por conseguinte, refletem seus praticantes, seguramente a IA pode e deve incluir emoções em seu repertório tecnológico. Esta linha produtiva segue: humanos socializam, humanos organizam-se em relações de poder, humanos percebem seus corpos... a todas estas provocações a IA responde com novas ideias para pesquisa e para construir sistemas *afetivos*, *sociais*, *políticos*... Podemos imaginar que, como o universo humano não é limitado em suas formas de *ser*, a IA não sofrerá com a falta de questões de pesquisa criativas.

O horizonte amplo de criações que replicam o espectro do *ser* humano, em conjunto com a possibilidade de atuar em projetos que procuram resolver, de maneira tecnológica, problemas práticos em áreas tais como educação, saúde ou segurança, tornam a IA atraente e instigante. Neste sentido, a IA é atraente e instigante também para o público:

histórias sobre suas realizações com frequência aparecem nas páginas de jornais, e há toda uma tradição de livros e filmes sobre o tema. Podemos mesmo dizer que é através desta tradição que podemos compreender a IA em uma perspectiva histórica, que transcende o momento que vivemos e se coloca em uma linhagem de desejos e receios do poder criador da prática técnica.

Minha aproximação com a IA ocorreu dentro deste panorama, que prometia desafio tecnológico e o acesso a conhecer e manipular múltiplas formas do *ser* humano. A situação era, mais precisamente, um curso em *computação afetiva*, que propõe tornar computadores capazes de lidar com a emoção. Isto envolve, segundo a computação afetiva, reconhecer as emoções das pessoas em seus encontros com a interface computacional, e projetar reações que estejam de acordo. Computadores sensíveis, as máquinas projetadas dentro deste paradigma deixariam para trás a insensibilidade de reagir sempre da mesma forma, repetindo vezes sem conta mensagens irritantes, e passariam a avaliar cuidadosamente as atividades propostas a seu usuário em função de como percebem seu estar afetivo.

O desenvolvimento do curso transportou os alunos da paisagem difusa e promissora das motivações da computação afetiva e das possibilidades de máquinas sensíveis a emoções para um terreno mais pragmático e próximo dos métodos para tratar a questão. A computação afetiva então começou a tomar forma, e a mostrar o trabalho exigido para concretizar suas propostas, assim como o cenário em que se desenvolviam a pesquisa e as aplicações desta técnica. Dentre os primeiros passos para dominar o projeto de sistemas dentro deste paradigma, estava justamente colocar o que aparentemente seria a noção central da computação afetiva: a *emoção*. A partir de um conjunto de trabalhos acadêmicos da disciplina da psicologia, um quadro de referência para compreender, categorizar, e estabelecer regras sobre a emoção foi construído para que dentro dele trabalhássemos.

Este momento marcou uma inflexão na minha percepção sobre as práticas da computação afetiva e da IA. O quadro referencial com o qual entrei em contato mostrou uma série de características que mostravam uma relação não casual entre si e com outras características da atividade tecnológica da computação em geral, e com a IA em particular. Um dos esquemas nocionais sobre emoção adotado pela computação afetiva,

em seus textos básicos, e também apresentada no curso, é a teoria psicológica das emoções chamada OCC (Ortony et al., 1990) – que será objeto de uma análise mais detalhada no capítulo 7. A razão dada para adotá-la foi a de que esta construção teórica presta-se bem para o emprego em projetos que envolvem a implementação de sistemas computacionais. De fato, esta teoria, que seus autores chamam de "Teoria Cognitiva das Emoções", propõe compreender as emoções dentro de uma estrutura bem organizada, em que emoções são consideradas como *estados internos*, bem caracterizados, discretos (isto é, separados e não contínuos), e em um número finito (na verdade, 22 estados diferentes). Estes estados correspondem à consequência de avaliações, julgamentos, a respeito de uma situação em que a pessoa – cuja estado emocional é o objeto de interrogação pela teoria – se encontra; e estas reações podem ser positivas ou negativas. Assim, as emoções são causadas pela avaliação em relação ao desenlace de eventos, ou a expectativa em relação a eventos, ou a ações realizadas por pessoas, e além do objeto de avaliação, diferem entre si (estruturalmente) em termos de valor positivo ou negativo, e (não estruturalmente) em termos de intensidade, e também por especificações suplementares mas não estruturais.

Ora, esta é uma noção um tanto peculiar de emoção, que chama a atenção por vários motivos, entre os quais a divisão clara do que é considerado como o espaço emotivo e a atribuição segura de uma avaliação moral, positiva ou negativa, ao indivíduo em seu encontro com o mundo. A preferência por esta forma de compreender a emoção é vinculada, pelos praticantes, à necessidade de "dispor de um modelo implementável", ou seja, de trabalhar a partir de uma interpretação do mundo que seja amena para os processos pelos quais artefatos computacionais são projetados e construídos; a teoria OCC, neste sentido, é considerada adequada pelos praticantes (sendo adotada em várias instâncias de projetos por este motivo, e sendo também divulgada dentro dos circuitos de aprendizagem do ofício).

Uma questão que solicita um exame mais detido começa a delinear-se, no entanto, se atentarmos para o fato de que os autores da teoria OCC, em questão, propõe como seu objetivo a construção de uma estrutura teórica *computacionalmente tratável*. Em outras palavras, a adequação da teoria à implementação mostra marcas de um movimento que não é apenas a escolha de um marco paradigmático entre outros, mas sim o enlace que por fim completa uma gênese de uma abordagem, a da *emoção computável*. A enunciação da emoção como fenômeno objetivo – destacado do corpo do sujeito, *dessubjetivada* – e

sua caracterização em categorias discretas e claramente positivas ou negativas começa então a encontrar um sentido mais amplo.

Passando do cenário da computação afetiva para o contexto mais amplo dos desenvolvimentos dentro da IA como um todo, o tema de *características humanas* como funcionalidades objetivas, adicionáveis a sistemas computacionais, mostra-se recorrente e produtivo. Além da emoção, são investigadas como objeto de enunciação computacional, por exemplo, a produção de sociabilidades, vínculos afetivos, narrativas, conhecimento de senso comum, cultura e rituais, e o ensino. Em outras palavras, a computação afetiva não está isolada em seu esforço. Cada uma destas abordagens parte de um conjunto de requisitos e premissas próprio da Inteligência Artificial, não necessariamente explícito em sua totalidade, e joga com maneiras de interpretar e compreender o *ser* e *agir* do humano.

A interpretação de características de ser humano pela IA já mereceu alguns exames, como o de Adam (Adam, 1998), em relação à sua postura de gênero subjacente, os de Collins (H. M. Collins, 1990) e Forsythe (Forsythe, 1993, 1999), sobre o que significa o conhecimento especialista nos sistemas, ou o de Suchman, que situa a noção de serviço embutida em propostas de sistemas inteligentes pessoais. Em comum a estes estudos, está a consideração da importância do tema, uma vez que sistemas de IA, ao serem trazidas para uso, carregam consigo capacidades políticas entretecidas a suas capacidades tecnológicas, e vão colocar as compreensões de humano, materializadas em sua construção, diante de pessoas que terão de lidar com estes sistemas. Estes estudos, apresentados na maioria durante os anos 1990, contribuíram significativamente para a colocação em perspectiva da produção da IA, mostrando como esta produz artefatos tecnológicos cuja eficácia é articulada com regimes de verdade que reivindicam universalidade e correspondência com o humano.

O panorama geral da informática computacional, e da IA especificamente, transformou-se dos anos 1990 para cá. A computação como intermediária de práticas cotidianas tornou-se mais ubíqua, seu papel para a circulação de informação tornou-se central com a difusão da internet, um amplo conjunto de novas avenidas de pesquisa foi buscado pela IA no que diz respeito a características do humano, e seus artefatos e técnicas estão difundidos e em contato com uma parcela muito maior de pessoas durante o dia a dia.

É neste cenário de inflexão que se situa a interrogação deste trabalho: *como são as noções construídas sobre o* ser *humano que são apropriadas com fins tecnológicos pelos praticantes da Inteligência Artificial*? Em outras palavras, como são escolhidos certos temas relacionados ao humano, e como são concomitantemente compreendidos e transformados em conceitos tecnologicamente operativos? A questão a ser investigada desdobra-se em outras, que se relacionam de maneira próxima e que contribuem para apreender a principal; o que caracteriza ou distingue a conduta propriamente humana, e a possibilidade de uma essencialidade humana para além da conduta, são problemas que surgem junto com aquele primeiro. A expressão destas noções não é necessariamente explícita, e em muitos casos está inscrita implicitamente no artefato, ou expressa nas expectativas sobre o encontro deste artefato com as pessoas. O posicionamento dos pesquisadores dentro dos universos sociais, isto é, como compreendem as possibilidades de circulação de seus artefatos nestes universos, também vêm compor o problema investigado nesta tese.

O problema colocado mostra relevância ao considerar-se a forma como a IA apropria e mobiliza de maneira peculiar conceitos sobre o humano, o social e o cultural, e os materializa em artefatos e expectativas de grande circulação e influência dentro da sociedade. A maneira peculiar a que nos referimos é fruto de desenvolvimentos teóricos e necessidades práticas da disciplina, e de uma negociação cuidados com vertentes de outras disciplinas, vertentes que são vistas como de certa forma compatíveis ou apropriáveis para os objetivos da IA.

A especificidade desta apropriação leva a formas de ver o humano, o social e o cultural que podem ser observadas a partir de perspectivas diferentes e oriundas de outras tradições intelectuais (como a dos Estudos de Ciência e Tecnologia, empregada nesta tese). A intenção é produzir um remapeamento de perspectivas, tornar visíveis e compreender confrontos que surgem na circulação desta tecnologia pela sociedade, e realizar uma crítica construtiva e contributiva para a própria prática corrente da Inteligência Artificial. Afinal, se os praticantes desta disciplina recorrem a variados constructos de outras disciplinas para constituir uma prática tão múltipla, o encontro com outros pontos de vista sobre esta prática também tem o potencial de ser produtivo para ela.

## *4.1 Universo de pesquisa*

Esta tese foi desenvolvida com base em trabalho de campo realizado junto a dois grupos de pesquisa em Inteligência Artificial, um dos quais é brasileiro, o outro sendo europeu (português). Estes grupos são de caráter acadêmico, e congregam, cada qual, professores de universidade, alunos de pós-graduação e outros pesquisadores. A pesquisa em ambos é orientada em torno de projetos relacionados à Informática na Educação ou de cunho interpretável como pedagógico. Seus projetos articulam, como mencionado na Introdução, através das habilidades e formação de seus participantes, várias técnicas e áreas da IA, tais como agentes artificiais, computação afetiva, IA simbólica, e até robótica, com o objetivo de criar sistemas *inteligentes*, muitos dos quais destinam-se à interação com pessoas, em processos educacionais ou de apoio à educação no sentido estrito ou de caráter pedagógico em um sentido mais amplo, tais como jogos sérios ou educativos. A identificação imprecisa dos grupos, assim como de seus participantes (que aqui serão referidos por pseudônimos), foi acordada, com os próprios participantes, como estratégia de preservação de suas individualidades.

A formação do grupo brasileiro é algo fluida; por este motivo, a delimitação que segui foi centrar minha atenção na professora Margarida, que é a líder do grupo, e nas pessoas, professores e alunos de pós-graduação, próximos a ela e empenhados em projetos que se relacionavam mais proximamente entre si e com Margarida. O grupo, que localiza-se dentro de uma universidade no Sul do Brasil, viceja dentro de dois cursos de pós-graduação da área de Informática que, embora próximos, são distintos em suas especialidades. Constitui-se em torno de "projetos", que são empreendimentos mais ou menos delimitados, centrados na produção de um artefato, que pode ser um sistema computacional funcional, mas que também pode ser um texto normativo ou metodológico sobre um problema específico. Através destes projetos é que são congregados, a partir da prof. Margarida, pesquisadores ao grupo, em função tanto de suas especialidades como também da proximidade dentro da universidade (proximidade diríamos geográfica) e também de trajetórias acadêmicas (uma proximidade social, dentro do terreno acadêmico). Os professores que participam em projetos dentro do grupo também possuem seus próprios grupos, em que desenvolvem outros projetos; observei em várias ocasiões estes outros grupos serem trazidos para junto do grupo de Margarida para compor uma

super-equipe com o objetivo de dar conta de um projeto de porte maior, que necessitava de um número maior de participantes – e que, dentro da dinâmica própria do universo acadêmico, também propiciava um maior aporte de recursos para os participantes.

Minha aproximação com este grupo ocorreu através de contatos com alguns dos participantes: pesquisadores que são alunos de pós-graduação, desenvolvendo seus próprios projetos de pesquisa, e professores que levam adiante projetos que alinham-se ou estão aninhados dentro dos projetos maiores do grupo. Apresentei as linhas de meu projeto, incluindo a justificativa para participar de seus momentos de trabalho, várias vezes, a cada pequeno conjunto de pesquisadores com que negociei minha participação. Uma dificuldade metodológica clara colocou-se desde o princípio, que se relaciona com o fato da minha formação enquanto engenheiro e pesquisador ser próxima à dos participantes do grupo. A consequência é que, perante minha perspectiva como observador, o cotidiano do grupo apareceu frequentemente como *inespecífico* ou em outras palavras *normal*: suas práticas são constituídas de atividades que cedem com dificuldade à observação, porque são próximas àquelas que realizei durante muito tempo e que considero simplesmente adequadas para dar conta dos objetivos do grupo.

Nesta fase da pesquisa, entrei em contato com uma candidata ao doutorado, Ariane. Ariane tinha a professora Margarida como orientadora, e estava no estágio intermediário de seu doutorado. Já tinha uma ideia do que desejava desenvolver, que seria um sistema computacional baseado em Inteligência Artificial para auxílio a aprendizado em uma situação específica, mas a definição de qual situação, quais estratégias computacionais seriam utilizadas, e quais bases teóricas seriam de interesse, ainda estavam no terreno das buscas. Algo já estava definido, no entanto: o trabalho seria articulado em proximidade com uma amiga de Ariane, estudante de mestrado na universidade do norte do país, de onde Ariane provinha. Esta situação levou a uma sistemática de trabalho, entre Ariane e sua companheira de pesquisa, que foi importante para abrir as portas "práticas" do universo do trabalho dentro de IA. Ambas as pesquisadoras comunicavam-se regularmente, quase todas as semanas, em longas sessões de mensagem instantânea na internet, para as quais passei a ser convidado. Nestas sessões tomei contato com questões de ordem bem pragmática, que remetiam à formulação gradativa do problema que se propunham a resolver, como colocar esta formulação em termos de requisitos para um possível sistema computacional, e como realizar esta construção ao longo de linhas

teóricas que dessem conta, isto é, justificassem, suas escolhas práticas e simultaneamente as informassem e orientassem. Assim, pude observar a proposição de soluções possíveis em termos de ferramentas computacionais (e teóricas) disponíveis, e as difíceis decisões sobre a conveniência ou possibilidade de utilizar uma ou outra ("Ontologias não fecha com usar agentes"), e, de muito interesse para esta tese, a procura de articulações com perspectivas teóricas que, estando fora do escopo imediato da IA, pudessem orientar a concepção do sistema no que diz respeito à sua relação – interação – com seus usuários putativos. Mais especificamente, tratava-se da Teoria da Relevância, da linguística, que relaciona a significância de enunciados para enunciador e para o destinatário. Ariane encontrou um professor em outra universidade local que é especialista nesta área, e teve a oportunidade de encontrar-se com ele para discutir a aplicação desta teoria ao seu trabalho.

Ao longo do tempo em que acompanhei o desenrolar desta pesquisa, cerca de um ano e meio, meu conhecimento sobre como um grupo de pesquisa em IA funciona foi sendo construído com a participação em outros momentos importantes. O acesso ao material pedagógico no ambiente online de uma disciplina básica de IA do curso de pós-graduação, lecionado pela prof. Margarida, colocou-me em contato com um conjunto considerado fundante de problemas, conceitos e autores da área disciplinar. Participei também de reuniões mais amplas, e mais formais, do grupo como um todo. Algumas destas eram de caráter organizativo ou administrativo referente a um projeto, e em que discutia-se o andamento dos trabalhos, a articulação dos resultados até então obtidos, e a distribuição de novas tarefas. Outras destas reuniões eram mais voltadas à pesquisa, em que participantes apresentavam, para uma audiência de professores do grupo ou próximos ao grupo (como acentuei acima, a distinção não é absoluta), o momento de sua pesquisa, discutindo princípios teóricos e resultados do trabalho. Todas estas reuniões "mais amplas" a que assisti caracterizavam-se pela presença da prof. Margarida; em reuniões mais administrativas sua condução da pauta era mais visível, enquanto que nas reuniões voltadas para temas mais relacionados a pesquisa sua liderança mostrava-se mais discreta, mais implícita, na medida em que temas e projetos em apresentação não eram apresentados como diretamente articulados a um projeto único, mas sim distribuídos, em termos de responsabilidade, pelos participantes professores orientadores. Acompanhei também o grupo de pesquisa, em estágio de formação, liderado por uma professora de

outra universidade, que era próxima e participante em certos momentos do grupo de Margarida.

Durante os últimos meses em que estive em contato com o grupo, tive a oportunidade de negociar minha participação em algumas das reuniões de trabalho de um outro projeto interessante. Lena é professora, participa de maneira muito próxima do grupo de pesquisa da prof. Margarida, e havia recém iniciado um projeto próprio de pesquisa em Computação Afetiva. O momento de início deste projeto foi muito oportuno, já que Lena teve a chance de formar o projeto com duas alunas recém-entradas na pós-graduação. Uma destas alunas já estava em contato com Lena há algum tempo, porque havia cursado disciplinas deste pós-graduação (como aluna externa) antes de ingressar, e sua formação é da área da Computação. A outra doutoranda, logo ao entrar, foi direcionada para a prof. Lena em função da proximidade de interesses em pesquisa, e sua formação é em Psicologia. O projeto consistia na aplicação da Computação Afetiva a sistemas informatizados de ensino, com a intenção de aperfeiçoar a interação do sistema com o aluno e também de prover o professor responsável com informação considerada relevante sobre o aluno e sobre a interação do aluno com o sistema. O que há de interessante neste projeto é a convergência entre a interdisciplinaridade solicitada pelo tema do projeto e a constituição do (pequeno) grupo por ele responsável: uma das participantes não apenas era especialista em área disciplinar que dialogava diretamente com a orientação do trabalho – isto é, em Psicologia, como dando conta em âmbito temático da *emoção* neste caso particular – como sua posição particular em relação ao tema, que encontrou-se bem com a abordagem do tema peculiar à Computação Afetiva. Por este motivo, de maneira semelhante mas complementar ao projeto levado adiante por Ariane, a minha participação foi frutífera ao acompanhar em poucas semanas o estabelecimento de um panorama para resolver na prática a seleção e implementação de soluções que, na bibliografia de referência, são considerados importantes.

Após este período de observação, a etapa seguinte iniciou, um pouco mais de meio ano depois, com o trabalho de campo em um outro grupo, de características (e orientação teórica) semelhantes, mas desta vez em um contexto diferente: na Europa. Ali, a inserção foi mais imediata, ao invés de gradual. O contato foi estabelecido com a líder do grupo, prof. Celeste, contando com o auxílio e o apoio de Margarida – o que facilitou esta etapa, já que ambas haviam sido orientandas do mesmo pesquisador português, durante a

década de 1990. Apresentei-me e ao meu projeto ao grupo como um todo durante uma das reuniões semanais que realizam, logo nas primeiras semanas após minha chegada a Portugal. Este grupo funciona de maneira mais compacta, mantendo uma estreita coesão no funcionamento cotidiano. Um dos fatores é a localização física: assim como o grupo brasileiro, este é formado por professores, alunos de pós-graduação e pesquisadores, mas ao contrário do grupo brasileiro, dispõe de uma área física própria. A área é um núcleo de salas concentrado em torno de um corredor ou ala de um grande prédio de um instituto de pesquisas tecnológicas, ligado a faculdades técnicas (especificamente o Instituto Superior Técnico e Instituto de Engenharia de Sistemas e Computadores). Este instituto de pesquisas está em um parque tecnológico a alguma distância de Lisboa (30 a 40min de automóvel), e desta maneira concentra o convívio dos pesquisadores durante o seu dia a dia, distanciado do movimento da metrópole. Fui designado para uma sala com dois colegas que iniciavam o doutorado, próximo às salas onde trabalhavam os outros alunos, pesquisadores e professores.

A proposta para realizar o trabalho de campo envolvia participar em uma atividade de projeto, dentro do grupo. Celeste, levando esta necessidade em conta, colocou-me em contato com Gepê, aluno de mestrado que estava entrando na fase de escrita de sua dissertação (lá, chamada de tese). Sua pesquisa envolveu a criação de um *agente* (uma classe específica de sistema computacional de inteligência artificial) que podia interagir com usuários a partir de duas plataformas distintas, um robô e um telefone celular. O sistema despertou meu interesse, em função do aspecto da programação envolvida e do resultado em termos de sistema interativo: era possível jogar xadrez com o robô, que comentava os lances realizados – e, sob um comando da interface, o robô entrava em "hibernação" e o telefone celular passava então a apresentar, em sua tela gráfica, a continuação do jogo, com um pequeno desenho animado do agente e seus comentários falados. Nem tudo correra como esperado em seu projeto, no entanto, e os dados estatísticos que Gepê coletara junto a usuários não haviam sido satisfatórios, mostrando-se (muito) inconclusivos após a análise. Esta situação, por sua vez, despertou ainda mais meu interesse, pelo contraste entre o produto de engenharia bem acabado com que eu interagira e a resposta problemática de seus usuários; após um período de amadurecimento, pudemos analisar em conjunto o experimento (análise esta que é o objeto do capítulo 5). Entrementes, ficou estabelecido que construiríamos um novo

sistema a partir do existente, e se possível realizaríamos testes de interação deste sistema com usuários. Fiquei encarregado de reprogramar a parte do telefone celular, tarefa com que fiquei envolvido durante os meses seguintes, enquanto que no processo aprendia a utilizar a plataforma de desenvolvimento e a dominar as interfaces de programação (*API*s) das bibliotecas básicas e de inteligência artificial utilizadas pelo grupo.

Em função da organização peculiar do próprio grupo, minha experiência de campo foi mais coesa em comparação com a brasileira. A convivência diária estava constituída no próprio decorrer de atividades cotidianas realizadas em conjunto, desde a paulatina chegada dos ocupantes às suas salas, pela manhã, passando pelo almoço em conjunto com os colegas alunos em algum restaurante próximo, até as tardes que alongavam-se à medida que o verão europeu chegava e que eram marcadas por pausas para o café, junto às máquinas de venda no andar térreo do prédio. O trabalho mais técnico que eu desenvolvia, entremeado com discussões com os colegas sobre temas relacionados, era o pano de fundo do dia a dia, pontuado por estes momentos "sociais".

Ao longo deste período, as semanas eram pontuadas por reuniões do grupo, sempre às terças-feiras pela manhã. A organização das reuniões era extremamente metódica, e estas constituíam-se em duas partes. A primeira parte era uma apresentação de "pesquisa" ou algum tema de trabalho, designada para algum participante ou convidado, seguida de uma arguição aguda e interessada por parte da audiência. A segunda, conduzida objetivamente pela prof. Celeste, era mais curta e com um andamento um pouco mais urgente (aproximava-se a hora do almoço), e tratava de assuntos de organização do grupo, tais como os prazos para entregar o material para a página do grupo ou os procedimentos para organizar a ida de algum participante a determinada conferência científica.

Novamente, como no caso com o grupo brasileiro, a participação no cotidiano de atividades tornou possível a aproximação com os temas, pontos de vista e valores considerados importantes para o trabalho do grupo em IA. A construção deste panorama não seguiu um processo explícito, já que alguns destes pontos surgem apenas como reação a situações específicas. Desta maneira, foi possível observar a importância dada às práticas que negociam o espaço do grupo dentro da comunidade acadêmica – tais como a atenção ao desenvolvimento da performance dos participantes na apresentação de suas pesquisas, e a disputa pela participação em projetos de grande porte financiados pela

União Europeia. No trabalho de pesquisa, observei o destaque dado implementação objetiva de sistemas e em testes de interação destes com pessoas; esta objetividade, como veremos, passava pela construção de modelos de características humanas com ênfase na tratabilidade computacional.

O período de campo em Portugal chegou ao seu fim em agosto, momento de férias acadêmicas no calendário acadêmico. A finalização do campo transcorreu de maneira que poderia chamar de anticlimática, já que à medida que decorriam as últimas semanas de julho e o mês de agosto, o número de pessoas no prédio foi diminuindo, em conjunto com as evasões também no pessoal do grupo. Com todos ansiosos pelo descanso anual, o movimento no corredor do grupo reduziu-se pouco a pouco, até o ponto em que mesmo a cantina do prédio, onde os remanescentes reuniam-se para o lanche, fechou. Nos últimos dias, alternávamos provas de interação do sistema que havíamos, eu e Gepê, construído, com partidas de jogos de tabuleiro entre os colegas nos momentos em que não havia eletricidade – o período de férias era usado para reformas no prédio, e a rede elétrica era desligada com frequência.

## *4.2 Procedimento da pesquisa*

O trabalho de realizar uma análise de um projeto tecnocientífico do tipo da IA, a partir de perspectivas sociais e culturais, demanda algumas soluções metodológicas não triviais. Buscou-se uma abordagem que fosse de fato interdisciplinar, isto é, o objetivo não era apenas o universo *social* da IA, ou sua produção *enquanto cultura* apenas, e sim uma visão que fosse inclusiva, e que desse nota do que conta como técnico, tornando visível sua imbricação em um contexto próprio, e que, ademais, situasse este contexto, examinando suas fronteiras e suas alegações de universalidade – próprias dos empreendimentos técnico-científicos.

Dada a natureza ramificada da IA, a qual dialoga com um grande número de outras disciplinas, dentre elas técnicas, científicas exatas, científicas sociais, e também a filosofia, encontrar uma perspectiva externa que dê conta de maneira monolítica e consistente não é uma tarefa fácil, e possivelmente não viável. Uma aproximação interdisciplinar, que

possa fazer conversar de maneira coerente perspectivas alternativas enquanto mantendo uma cesura clara, isto é, buscando manter um espaço analítico que não seja subsumido pela própria abundância do conhecimento da IA, parece-nos portanto a melhor opção.

Dito isso, o próprio panorama dos estudos de Ciência, Tecnologia e Sociedade (*Science and Technology Studies*, STS), área interdisciplinar que tem se proposto a investigar a produção da ciência e da tecnologia, é informativo em relação à necessidade de métodos peculiares para lidar com o tema. O ponto de partida usualmente é o da observação do *processo* da produção na ciência e também na técnica. O conhecimento científico estabilizado é apoiado em uma densa rede de interconexões e consensos, ou, como também é dito, em *verdades científicas*. Da mesma forma, tecnologias estabelecidas apenas *funcionam*, tendo nelas sido apagada a trajetória contingente que a tornou uma escolha hegemônica. Tal rede é de difícil análise, não oferecendo pontos acessíveis para exame que transcenda a verdade ali acomodada. Por conseguinte, o melhor caminho a tomar é levar em conta o *processo*, o conhecimento em formação, não estabilizado e mostrando a marca da intervenção criativa e das decisões experimentais e da negociação entre os envolvidos na busca pela resposta ao problema proposto.

Em outras palavras, uma das dificuldades para o presente trabalho é seu caráter reflexivo: consiste em aplicar métodos de investigação científica à produção de conhecimento legitimada *dentro* da ciência. Não é apenas a descrição desta produção, nem tampouco uma investigação epistemológica circunscrita no sentido de estabelecer ou negar reivindicações de verdade. Pontos epistemológicos estão presentes, e serão discutidos no devido tempo, mas o objetivo mais amplo é problematizar a IA como uma forma de produção de conhecimento, como uma forma de produção cultural, como uma forma, material e conceitual, de produzir formas de *viver* e de realizar agenciamentos, imbricados a determinados objetos técnicos, em nossa sociedade.

Um dos métodos utilizados com alguma frequência por projetos de pesquisa em CTS é a etnografia (Callon, 1986; Guimarães Jr., 2005; Hess, 2007; Latour, 2000; Suchman, 2000). A etnografia enquanto método apresenta algumas características apropriadas para a tarefa, uma vez que seu objetivo é desenvolver conhecimento e pontos de vista sobre uma determinada cultura, através da prática da observação participante. A observação participante, central para a etnografia, significa o envolvimento do pesquisador em algum

papel participativo, dentro do ambiente estudado, em sua configuração cotidiana, usual, "natural". Outras características são a abordagem integral, isto é, aspectos da cultura são considerados como relacionados ao todo, a sensibilidade ao contexto, ou seja, a busca de considerar o contexto como parte da explicação de suas observações, como um todo experiencial, e ainda a descrição em termos de sociedade e cultura em relação entre si (STEWART, 1998). Desta forma, a produção de ciência e de tecnologia é observada de uma maneira abrangente, em que as ações e as expectativas culturais de seus praticantes são considerados; justamente os detalhes que a prática usual e os valores da comunidade científica procuram alijar do seu produto final, na medida em que reivindicam para este universalidade e objetividade.

Seguindo esta orientação, o presente trabalho procurou explorar detalhadamente o fenômeno social em questão, em um pequeno número de casos, a partir de dados que, em um primeiro momento, ainda não estão estruturados em termos de categorias analíticas. Procura-se interpretar os significados e as práticas observadas, constituindo uma visão a partir de dentro do grupo: o objetivo é compreender o conhecimento do participante, enquanto pertencente a uma cultura própria e peculiar, e a lógica interna desta cultura. Dentro desta estratégia etnográfica, a observação participante (FLICK, 2004; J. LOFLAND et al., 2005), como uma das práticas centrais de construção de dados, procura acessar as práticas conforme o universo pesquisado as encena, exigindo que o observador exerça suas competências observacionais de uma maneira integral em relação ao ambiente em que decorre a ação. A motivação é obter conhecimento a partir das práticas, não apenas a partir do relato das práticas. Dentro do contexto específico do universo explorado neste trabalho, um grupo pequeno de pesquisa acadêmica, o observador também é participante: a sua presença é conspícua, e o observador precisa por este motivo encontrar e negociar um papel que permita sua presença e que, além disso, permita que as práticas próprias do grupo sejam encenadas mesmo com sua presença (em outras palavras, que o trabalho do grupo possa ser realizado da forma mais usual possível). Esta situação é necessária até por motivos pragmáticos, uma vez que o observador precisa realizar sua pesquisa em um prazo longo mas ao mesmo tempo o grupo precisa levar suas próprias tarefas a cabo.

O processo de pesquisa seguiu atentamente uma série de orientações próprias do método etnográfico, em função das necessidades para a construção dos dados na relação com o sujeito particular de pesquisa, isto é, o próprio universo de pesquisa científica e

tecnológica. Em conjunto com a observação participante nos grupos, outros registros da realidade foram investigados, principalmente documentos escritos, que constituem uma das formas de produção mais valorizados dentro do contexto acadêmico em que estes grupos atuam. Artigos e relatórios condensam, em uma forma considerada adequada para apresentação aos pares, tanto as premissas de prática e conhecimento como a apreciação dos resultados observados; neste sentido, são fontes importantes para entender como os casos particulares observados são apreciados e avaliados pelos cientistas envolvidos.

Além da produção escrita, foram interrogados com especial atenção os artefatos produzidos: sistemas computacionais, investidos das características funcionais desejadas pelos participantes, e materializando suas práticas e premissas. Em vários momentos, e de maneira não uniforme, foi possível entrar em contato com estas corporeo-realizações para dentro da experiência do trabalho de campo. Este contato não foi sistematicamente uniforme porque, além da natureza construtiva, a maneira como estes artefatos tornam-se disponíveis tampouco é "uniforme" para quem está em campo. Os sistemas computacionais em questão "funcionam" quando uma série de pré-requisitos operacionais está colocada, como uma infraestrutura que os sustenta. Precisam ser instalados, inicializados; o equipamento que é sua base precisa estar disponível, o que não é sempre o caso quando estamos considerando equipamentos mais específicos, e mais cuidadosamente cuidados pelos participantes (como é o caso do robô iCat descrito no capítulo 6). A interação com estes objetos também deve ser objeto de cuidadosa atenção, e o resultado é que não ocorrem muitos momentos em que podem ser observadas a interação com pessoas, e muito mais raros os momentos em que esta interação é com pessoas de fora dos grupos de pesquisa. Alguns destes objetos são disponíveis como vídeos gravados, em que certos intervalos em que o sistema está funcionando grava-se o resultado apresentado na tela; outros, com um pouco de sorte, estão disponíveis como pacotes de software que podem ser baixados da internet e, através de algum processo mais ou menos trabalhoso, podem ser feitos funcionar em um computador, para que possam ser observados.

O processo de construção, através das habilidades que devem ser dominadas, da decisão sobre requisitos, da deferência dada aos referenciais considerados mais ou menos canônicos, e de como este processo é acompanhado e avaliado dentro do grupo, também fez parte da constituição da pesquisa. Este é um processo curiosamente "invisível", na

medida em que ocorre ao longo de um tempo longo, e possui um grande número de inespecificidades para a observação participante. O processo de formação técnica é muito longo, e observá-lo acontecendo, sendo um "nativo operativo" como é o meu caso, invisibiliza grande parte das premissas que são trazidas já construídas dos processos anteriores; por outro lado, um estranhamento radical, como o de alguém não formado dentro das competências técnicas correntes no universo de pesquisa, seria uma barreira para o mergulho dentro das práticas cotidianas que ali ocorrem – tais como a existência de programas de computador considerados "disponíveis" mas que demandam um processo de instalação trabalhoso, ou a constituição de um ambiente de trabalho, no computador pessoal, a partir de um grande número de ferramentas computacionais que devem ser colocadas em operação para possibilitar o trabalho cotidiano da programação do sistema. A realização do presente trabalho deriva de uma negociação entre estes extremos.

No entanto, esta atenção a detalhes que ligam as práticas, significados e artefatos não significa que o resultado seja uma "etnografia" no sentido estrito como as da tradição de pesquisa antropológica ou como as das orientações metodológicas em Stewart (1998). Esta situação ocorreu, no decurso deste trabalho, por conta de várias razões.

As razões são ligadas à gama de dados e o tipo de análise que se pretendeu constituir, em função da forma como o trabalho objetiva contribuir para a área da Informática na Educação. A análise empreendida procura constituir um trajeto que perpassa o trabalho dos grupos acompanhados, e que, para tornar mais claros os pontos colocados, estende-se em direção ao campo mais amplo a que se vinculam os grupos, e, em direção contrária, para fora do grupo, próximo a quem utiliza os produtos computacionais, em um outro momento. Este último caso constitui-se da análise do encontro de um produto computacional de um dos grupos com um conjunto de usuários, isto é, pessoas fora do grupo. O anteriormente mencionado é o encontro da trajetória de pesquisa aqui relatada com um sistema em particular, com um objetivo que remete a discussões epistemológicas, que é o do conhecimento de senso comum, e que – por este motivo – é reconhecido e citado como exemplo, dentro do campo e dos grupos.

Há uma outra razão importante para o trabalho ter derivado sua estratégia de construção em uma direção que de certa forma se afasta do que seria um resultado etnográfico. O trabalho etnográfico decorre da aplicação de um saber específico, isto é, é

um trabalho especialista, e adquirir uma performance considerada adequada dentro desta prática demanda tempo e experiência. Embora eu tenha desenvolvido alguma proficiência em técnicas relacionadas a este método, constituir o texto inteiramente dentro dos cânones etnográficos não seria possível dentro dos limites de recursos e tempo disponíveis. A solução foi um compromisso que trouxesse para o trabalho as dimensões consideradas essenciais, como a contextualização e a construção de conhecimento a partir da observação participante.

Por fim, é necessário apresentar uma característica deste trabalho que é trazida do tipo de análise sistematicamente utilizada na disciplina que desenvolveu o método etnográfico, que é a Antropologia. Procuramos consistentemente trabalhar o "colocar em perspectiva" das práticas e noções próprias do campo estudado. As práticas e noções, no campo abordado, são conhecimento científico e tecnológico, com estatuto próprio e estabelecido como válido e legítimo; a análise não coloca isto em questão, nem procura desvalidar os processos utilizados. O que sistematicamente foi buscado foi a colocação desta validade e desta legitimidade *em perspectiva*, mostrando como esta validade relaciona-se com a forma de produção e legitimação, e como esta produção e legitimação podem ser vistas de outra forma. A intenção é conduzir o debate para a consideração de outras maneiras de perceber a legitimidade, e de produzir tecnologias e formas de viver nas quais pontos de vista alternativos tenham sido avaliados, não apenas os da ciência, ou os da Inteligência Artificial.

# 5 Inteligência Artificial: especialistas e a demarcação dos territórios do saber[6]

Sistemas Especialistas, também chamados de sistemas baseados em conhecimento, são programas de computador que codificam em uma *base de conhecimento* conjuntos de conhecimento, como fatos, relações e regras, de disciplinas específicas (ver capítulo 3). O presente capítulo procura investigar, através dos sistemas especialistas, uma das formas de ser "*ser humano*" de que a IA se apropria e coloca, exercendo-a no domínio da máquina, ao serviço das pessoas: o conhecimento. Abordaremos um sistema especialista em particular, Cyc (Guha & Douglas Lenat, 1990), que procura diferenciar-se de outros sistemas através da codificação, em sua base de conhecimento, do *senso comum*. A proposta de seus autores é criar um sistema capaz de lidar com um grande número de situações simples nas quais outros sistemas falham porque são restritos em seu escopo.

A problemática da codificação e representação do conhecimento aparece com frequência quando são abordadas, dentro dos grupos em que foi realizado o trabalho de campo, as formas de trabalho e os objetivos da Inteligência Artificial. Sistemas especialistas são clássicos exemplos para o tema, porque estão estabelecidos já há longo tempo enquanto solução de engenharia para determinados problemas práticos. No entanto, são clássicos também no sentido de que funcionam melhor para corpos de conhecimento específicos, especializados e altamente formalizados, como por exemplo especialidades acadêmicas ou profissionais. Cyc, em discussões sobre este tema, é citado como um esforço especial para produzir um sistema que opere, através das técnicas de sistemas especialistas, além destas fronteiras, no terreno do conhecimento de senso comum. O escopo de Cyc pretende ser "amplo", incluindo todo o conhecimento

6 Uma versão do material deste capítulo foi publicada como (Wild et al., 2011)

corriqueiro que as pessoas empregam comumente para realizar as tarefas de seu dia-a-dia. Os autores do sistema descrevem de várias maneiras o que seria "todo" esse conhecimento, "comum" a todos e "corriqueiro". Por este motivo, trata-se de um contexto adequado para analisar o quê os proponentes da IA entendem por conhecimento, e como manipulam e colocam em uso este conceito.

Procuraremos examinar como esse conhecimento é descrito e como é produzido. A partir de exemplos provenientes de um conjunto de textos produzidos pelos criadores de Cyc, e também de sessões de interação com o sistema, buscaremos questionar algumas das premissas implícitas na proposta de Cyc e colocar em perspectiva as características que seus autores expressam sobre esse conhecimento. Essas características declaradas para Cyc, tais como de ser um conhecimento universal e consensual, acessível e compreensível, podem ser interrogadas, com o intuito de mostrar como estão relacionadas à adoção de pressupostos epistêmicos que apagam seu caráter contextual e implicado com o sujeito. A partir destes questionamentos, nossa intenção é tornar visíveis as condições em que este conhecimento se torna *válido*, e como é ancorado em um regime de verdades que provê para sua validade e que, ao invés de universal e consensual, é situado e contingente.

## *5.1 Conhecimento e Sistemas Especialistas*

Sistemas especialistas são interessantes porque materializam e colocam em operação uma interpretação específica do que é *conhecimento*. A construção e a utilização de bases de conhecimento em sistemas especialistas tornam visível a abordagem epistemológica que as suporta. Conhecimento, para os proponentes desses sistemas, é visto como um conjunto de afirmações e regras declarativas sobre um dado campo de saber. Trata-se, sob este ponto de vista, de um conjunto objetivo, coerente e não-problemático, sob domínio da pessoa que é especialista na área. Essa posição é desdobrada de várias formas: na maneira como é afirmada a *existência* de um tal conhecimento, na abordagem do problema como sendo uma questão de *representação* do conhecimento, e na pouca atenção que é dada aos processos pelos quais esse conhecimento é obtido – a partir do especialista, pelo engenheiro de conhecimento.

Para compreender melhor as questões que procuraremos colocar, é importante prestar atenção ao conceito de conhecimento que é empregado em cada caso, e em como este conceito é colocado em prática. Harry Collins (H. M. COLLINS, 1990) examina detidamente os "sistemas especialistas" e a maneira como *inteligência* é atribuída a esse tipo de dispositivo tecnológico. Partindo de uma perspectiva de estudo cujo objeto é a tecnologia e a ciência, Collins mostra como sistemas especialistas inserem-se em uma tradição mais ampla de desenvolvimento de máquinas que são construídas para executar, de forma codificada, ações específicas que são parte importante de tarefas que envolvem práticas cognitivas e culturais. Nessa perspectiva, as máquinas *funcionam* quando as ações codificadas que são capazes de realizar são embutidas dentro da prática mais ampla das pessoas que utilizam e interpretam essas ações codificadas. Um exemplo apresentado por Collins é a máquina calculadora simples, que codifica a ação integrada de teclas e visor que, quando operada e interpretada por uma pessoa com a destreza necessária, realiza a função aritmética de somar ou multiplicar.

Collins mostra que esse também é o caso do sistema especialista que representa o conhecimento em uma base digital de conhecimento e de regras. Esses sistemas são cuidadosamente construídos para codificar o trabalho complexo de um especialista em uma determinada área de atuação. Como resultado, a própria operação do programa exige familiaridade com a área, já que é necessário conhecer os conceitos em cujos termos serão formuladas as interações com o programa, colocar o problema e compreender a resposta obtida, e, em resumo, interpretar no mundo real o que conta como atualização válida dos conceitos, critérios e resultados da área. Em outras palavras, se um sistema especialista de apoio ao diagnóstico médico menciona "febre", o operador do programa precisa competentemente interpretar esse conceito no contexto do corpo do paciente, sabendo pragmaticamente como medir febre – manipulando o corpo do paciente e o instrumento de medição – e como decidir sobre a categoria "febre" a partir dos resultados da prática – qual foi a temperatura medida, em qual parte do corpo do paciente, em qual horário do dia.

O conhecimento codificado e enunciado no computador, portanto, não funciona isolado e independente em um mundo conceitual, mas embutido em um universo de práticas culturais de pessoas que o interpretam, o utilizam e assim o tornam significativo. *Conhecimento é interpretado* e não é auto-evidente; o significado de um enunciado

particular do conhecimento é construído dentro do compartilhamento de um conjunto de conhecimentos sobre o mundo (FORSYTHE, 1993).

Essa é uma forma de compreender uma das características relevantes dos sistemas especialistas: a necessidade de que a pessoa que os utiliza conheça também os conceitos e processos envolvidos na prática da especialidade à qual o sistema é dedicado. É necessário interpretar adequadamente os passos de interação do sistema e saber manipular a informação solicitada e fornecida. Em suma, é preciso que o sistema seja operado por um humano que entenda o que está sendo processado e possa compreender a resposta e colocá-la em prática.

Além disso, esses sistemas lidam com dificuldade com situações envolvendo conhecimentos que não estão explicitamente incluídas no universo de conhecimento, mesmo que conhecimentos aparentemente simples e não especializados, com os quais pessoas estão acostumadas a lidar. Essa dificuldade pode ser sentida, por exemplo, em livros-texto, na cuidadosa apresentação de maneiras de efetuar, para efeitos de IA, categorização de objetos que no cotidiano são apreendidos de maneira prática por pessoas em seu dia-a-dia. Um exemplo muito citado é a categoria "ave"; aves voam e essa é uma característica importante desses animais, mas há aves que não voam, tais como avestruz. Outro exemplo é o de conclusões implícitas relacionadas a certos fatos: pessoas possuem normalmente dois braços, entretanto há pessoas que possuem apenas um braço; ademais, pessoas que possuem apenas um braço possuem apenas uma mão (ambos exemplos em (HARMON, 1985, p. 37). Conta-se (FORSYTHE, 1993; LUGER, 2002) que o sistema MYCIN, certa vez, teria sugerido gravidez como causa do estado febril de um paciente do sexo masculino. Embora não seja necessário ser um especialista médico para saber que apenas mulheres ficam grávidas, um sistema especialista não tem acesso a esse conhecimento a não ser que seja de alguma forma representado em sua base. Essa tendência a falhar quando a interação do sistema foge do escopo específico, mesmo que o que esteja em jogo não seja conhecimento especialista, é conhecida como *fragilidade* do sistema baseado em conhecimento (GOOGLE TECHTALKS, 2006; PANTON et al., 2006).

Essas características não costumam ser uma limitação problemática em muitos casos; o esforço de projeto para a criação desses sistemas é focalizado para áreas de saber altamente formalizadas e cujos problemas a resolver são bem definidos, com abrangência

muito específica, propositadamente deixando de lado a possibilidade de interação fora do escopo escolhido para o sistema. Além disso, espera-se que a operação desses sistemas, como já mencionado acima, seja realizada por pessoas que dominem a respectiva área de saber, e que portanto interpretam adequadamente os conceitos e procedimentos mencionados pelo sistema.

Mesmo que a restrição a escopos de aplicação específicos e a utilização por pessoal qualificado dêem conta das características apresentadas acima, persiste, entre os proponentes da Inteligência Artificial em geral, e dos sistemas especialistas em particular, a questão de como construir sistemas que não demonstrem a mencionada *fragilidade* em relação ao conhecimento não-especialista do mundo.

Dentro desta perspectiva, um relevante estudo da prática da Inteligência Artificial é o da antropóloga Diane Forsythe (Forsythe, 1993). Seu estudo surgiu a partir de um trabalho de campo que durou vários dentro de um conjunto de laboratórios acadêmicos de pesquisa em IA, os quais desenvolviam um tipo específico de sistema computacional: sistemas especialistas (ou sistemas baseados em conhecimento). A questão que Forsythe aborda é o conceito de "conhecimento" para os engenheiros dos laboratórios: o que o termo significa para estas pessoas, como esse significado se expressa no trabalho realizado, e as implicações políticas desta situação. O trabalho de campo da autora foi desenvolvido a partir de observação participante em laboratórios, e também de encontros, conferências e reuniões em que os pesquisadores do laboratório tomavam parte. Para atingir seus objetivos, Forsythe procurou investigar a relação entre as concepções e ideias que a comunidade de pesquisadores compartilhava, as práticas que constituíam seu trabalho, e as características dos artefatos produzidos por seu trabalho. Pertencer a esta comunidade, segundo observou Forsythe, implicava aceitar e compartilhar um certo conjunto de significados e práticas com os outros membros; ser reconhecido como compartilhando estes significados e práticas funcionava como um critério relevante para pertencer à comunidade. Uma parte importante do estudo constituiu-se na tentativa de compreender como os pesquisadores posicionam este saber próprio da área disciplinar em relação a contextos culturais mais amplos e genéricos dos quais também participam, mas em contraste com os quais definem-se como pesquisadores em IA. Isto é, segundo a autora, no movimento de construir artefatos e significados tecnológicos próprios de sua área, estes cientistas apoiam-se em um repertório de saberes e práticas familiares sobre o

ordenamento do mundo, alguns dos quais são explicitados, mas muitos dos quais são trazidos de forma tácita, não explicitada.

A partir destas observações, Forsythe situou o trabalho dos pesquisadores desenvolvedores de sistemas especialistas. Sistemas especialistas, segundo Forsythe, são projetados para "emular *expertise* humana" (FORSYTHE, 1993, p. 451). Para construí-los, é necessário coletar informação, ordená-la, e projetar um programa computacional para manipulá-la. *Expertise*, conhecimento especializado, é considerada uma forma particular de raciocínio e informação que um certo tipo de pessoas possui: os especialistas (ou *experts*). A questão que se coloca, para a realização prática destes sistemas, é o que os especialistas com quem Forsythe trabalhou chamavam de "aquisição do conhecimento". A autora contrasta o ponto de vista das ciências sociais com a abordagem que ela observou, nos laboratórios, em termos de engenharia. Do ponto de vista das ciências sociais, da antropologia em especial, obter "conhecimento" é uma tarefa complexa, na medida em que o engenheiro precisa compreender tanto o processo pelo o especialista chega às suas decisões, como também esta informação precisa ser colocada em linguagem acessível à máquina. Forsythe destaca algumas características do conhecimento, na perspectiva antropológica. Em primeiro lugar, ressalta o fato de ser socialmente e culturalmente constituído, isto é, não é simplesmente um decalque da natureza na linguagem humana, mas construído e mantido dentro do grupo que o constitui. Além disso, uma parte considerável do conhecimento não é explicitamente declarado, estando implícito nas práticas das pessoas ou distribuído em divisões do trabalho ou procedimentos cotidianos. Somando-se a isso, ocorre que a prática observada das pessoas não corresponde exatamente à prescrição para a tarefa, nem à descrição que fazem dela. Estes fatores, tomados em conjunto, indicam que "adquirir conhecimento" para um sistema especialista não é uma atividade simples.

Os cientistas do laboratório colocavam a questão de maneira diferente. Forsythe conta que, para estes cientistas, o problema era que os humanos que deveriam ser a "fonte" do conhecimento são ineficientes (por serem humanos), tornando o processo lento e tedioso. Na linguagem das ciências sociais, os cientistas do laboratório "reificavam" o conhecimento: referiam-se a ele como algo concreto, uma entidade bem delimitada que pode ser "adquirida" ou "revelada". Conhecimento, assim, era visto como uma questão

binária, ou presente ou ausente, ou correto ou errado, caracterizado como um fenômeno puramente cognitivo baseado em regras claras e conscientes de raciocínio.

Partindo destas considerações, Forsythe coloca uma série de questões que consideramos como de grande alcance para a discussão sobre Inteligência Artificial na relação com o social. Uma destas é, seguindo o conceito de apagamento como formulado por Star (Star & Strauss, 1999), a forma como a posição dos pesquisadores *apaga* a natureza social de seu próprio trabalho. As sessões de entrevistas com especialistas demandavam, dos pesquisadores delas encarregados, uma atividade de negociação com o especialista e de interpretação das respostas, para situá-las a partir do contexto do especialista e em direção ao contexto do sistema em desenvolvimento, e para adaptá-las a uma forma adequada a este sistema. Esta atividade, apesar de complexa e repleta de dificuldades, era vista de maneira simplificada e pouco elaborada pelos pesquisadores, descrita apenas como um processo de obter respostas a partir de perguntas. Em conclusão, os pesquisadores do laboratório *apagavam* de suas próprias atividades o quanto continham de social e de cultural.

## *5.2 Posição do sujeito no conhecer: uma posição política*

A questão do conhecimento como forma de *poder* também é colocada pela autora. Estar apto a decidir o que conta como "saber" é uma forma de poder, exercida sistematicamente pelos pesquisadores em sua tarefa de produzir sistemas computacionais baseados em conhecimento. A definição que propõem para "conhecimento" embute também, na realização destes sistemas, algumas características importantes, mas não explicitamente discutidas. Partimos aqui do pressuposto de que um corpo de conhecimento estabelecido traz consigo as marcas de quem o construiu e de sua proposta de mundo. O conhecimento dito *verdadeiro* é algo "criado" e reforçado por encontros e desencontros de forças, onde "quem" o afirma é, portanto, relevante na comunidade de observadores que compartilham essa evidência. Há um contraste dessa posição com a epistemologia tradicional do "sujeito universal", na qual é apagada a situação particular de *quem* enuncia o saber, como destacado por (Adam, 1998). Levamos, pois, em conta que

critérios de validação não estão referidos a uma determinada realidade independente, mas resultam de uma tensa consensualidade de um coletivo de pesquisadores que compartilham um domínio de práticas e conceitos. Essa consensualidade tece uma rede densa que estabiliza o conhecer, ao mesmo tempo tornando-o válido (e, portanto, produtivo) e marcando de forma peculiar essa validade. Um exemplo é o debate travado, na Inglaterra do século XVII, sobre o estatuto do *fato científico* experimental, no caso a bomba de vácuo inventada por Robert Boyle (Shapin, 1985). A questão, proposta por Boyle e que se tornou fundamental para o avanço da física experimental como a conhecemos, era como estabelecer fenômenos como *fatos*, dado o problema de quem havia encenado o experimento, como o havia feito, e quem eram as testemunhas – e, principalmente, dado o problema da existência de outras formas de se chegar à *verdade*. A proposta de Boyle entrava em conflito com outras categorias de verdade em uso e consideradas como válidas, como por exemplo o uso da razão como forma de validação de uma proposição. O conhecimento é resultado de um campo de lutas e contradições, em que a interação com outros domínios de saber e relações de poder entram em jogo para tornar alguns enunciados mais legítimos que outros. Nesse sentido, a verdade não pode ser tomada como algo universal e passa a ser compreendida também como uma questão política, enredada nos critérios de validação hegemônica (ver, por exemplo, a discussão em (Dreyfus, 1983, pp. 114-117)).

O conhecimento se produz como político nestes sistemas: diferenças em cultura, classe e gênero são apagadas, na medida em que a comunidade que constrói estes sistemas é muito homogênea. Este conhecimento, colocado como verdade, tem um efeito político, reiterando as condições que o tornaram possível. É importante, por este motivo, observar como relacionam-se entre si o aspectos técnicos explícitos e os aspectos políticos, menos explicitamente discutido, dentro dos projetos da IA. As características técnicas, descritas de uma maneira utilitária e funcional, ligam-se à proposta epistêmica em que baseiam-se, e por sua vez as soluções encontradas para o problema técnico vão fazer parte da constituição de suas formas políticas. Discutir, ou não, este conhecimento, depende de quem é enuncia e dá garantia a esse conhecimento, e o resultado é a legitimação do ponto de vista dos grupos que dominam esta construção. Assim estabelecem-se certas verdades e apagam-se outras, em processos que se materializam como ferramentas, revestidas da legitimidade simbólica da "Computação" e da "Inteligência Artificial"

## *5.3 Suprindo um déficit: Cyc*

Cyc (Guha & Douglas Lenat, 1990), portanto, é um dos projetos mais conhecidos por seu esforço em trazer o senso comum ao domínio dos sistemas especialistas. Douglas Lenat iniciou em 1984 o projeto, que continua em andamento sob sua coordenação. A abordagem de Cyc para a questão do senso comum é "prover um repositório de senso comum formalmente representado". Lenat propõe que um repositório desse tipo pode funcionar como uma espécie de substrato referencial ao qual sistemas especialistas "especialistas" poderiam recorrer quando o problema não está contido em seu domínio específico, mas não requer, por sua vez, recurso a uma outra especialidade. A motivação é apresentada por Lenat (Google TechTalks, 2006) com exemplos jocosos em que sistemas especialistas falham. Um destes exemplos é o de um sistema médico especialista, em que o "consultando" é um velho e enferrujado automóvel Pontiac com "manchas marrom-avermelhadas no corpo": o diagnóstico do sistema para o "paciente" é sarampo. Lenat aponta essa situação como um exemplo da mencionada fragilidade, e a considera como uma limitação e um problema a ser resolvido em um contexto em que aplicações de sistemas de Inteligência Artificial tornam-se mais difundidos e mais importantes.

O senso comum com o qual prover Cyc é, segundo seus proponentes, aquele "conhecimento geral que nos permite sobreviver no mundo real, e entender e reagir com flexibilidade diante de situações novas" (Panton et al., 2006) e constitui-se de fatos e regras empíricas e "assim por diante" que devem ser passadas para o sistema. O objetivo é coletar um número muito elevado de afirmações válidas sobre o mundo, um corpus de conhecimento que inclua o que pessoas sabem, e que permita ao sistema fazer inferências lógicas sobre questões "não especialistas", inferências que seriam as mesmas que as pessoas fariam sobre essas questões.

A colocação do conhecimento de senso comum como simples e não-discutível indica que esse deveria ser, de acordo com a proposta de Cyc, composto de declarações diretas, imediatas, sobre o objeto de conhecimento, sem margem a interpretações que, vistas como ambiguidades algo incômodas, obscureceriam a verdade sobre o objeto. Por outro lado, há evidências de que o conhecimento objetivo e compartilhado sobre o objeto, incluindo o contexto cotidiano do senso comum, é efeito do estabelecimento de uma teoria sobre esse

objeto. Essa teoria é construída pelo indivíduo a partir do corpus de conhecimento socialmente adquirido, e da interpretação da sua experiência sensória sobre o objeto. O objeto torna-se claro e manuseável quando manipulamos e recriamos as categorias e teorias de que dispomos para ali acomodar o novo objeto (KARMILOFF-SMITH & INHELDER, 1975) – isto é, ao relacionar de forma exploratória, repetidas vezes, o objeto com a teoria de que dispomos. É dessa maneira que distinguimos novos objetos, tais como novas frutas que conhecemos, e as acomodamos dentro do nosso universo alimentar, distinguindo-as de outros objetos que em princípio compreendemos diferentemente. Na teoria cotidiana, que eu enquanto sujeito conhecedor-em-senso-comum compartilho, maçãs são próximas de maracujás, e separadas de tomates, em razão da categoria fruta dentro das quais as recebo e interpreto (há categorias, por exemplo, em que o tomate é uma fruta).

Um extenso corpo de pesquisas, por exemplo as formuladas pelas etno-epistemologias e pela epistemologia enativa, sugere que a construção do entendimento do mundo não é consensual ou universal. Humberto Maturana (MATURANA & VARELA, 2001) coloca que o conhecimento está sempre relacionado "à operação de distinção de um observador". Assim, um conhecimento é para alguém e nunca em si. O observador não é preexistente, mas se constitui como tal inscrito corporalmente em uma comunidade de observadores que compartilham operações congruentes e critérios de validação. O conhecimento é então resultado de coordenações de ações em mundo. Em outras palavras, o conhecer não é uma questão representacional, mas uma agência inscrita no sujeito conhecedor através do viver. Por outro lado, a ação concertada e o compartilhamento de saberes são chaves pelas quais o indivíduo aprende a interpretar o mundo, e é o contexto em que é demonstrado como "válido".

## *5.4 Uma perspectiva particular sobre o universal*

Ao mesmo tempo em que a verdade se produz politicamente, o efeito dessa mesma verdade é político. Em geral, ela reproduz as relações que a tornaram possível. Nesse sentido, é interessante observar como dialogam entre si o aspecto técnico, explicitamente assumido, e o aspecto político, menos explicitamente discutido, dentro do projeto Cyc. As

características técnicas desejadas para Cyc estão ligadas ao projeto epistêmico que lhe dá sustentação, e por sua vez as soluções encontradas para o problema técnico vão fazer parte da constituição de suas formas políticas. Colocar certos itens de conhecimento em discussão, ou não, depende de *quem* é a fonte desse conhecimento, e o resultado é a legitimação do ponto de vista dos grupos que participaram nessa construção. Essa é uma forma de estabelecer certas verdades e apagar outras, materializada em uma ferramenta que propõe, revestida do prestígio simbólico da "Computação" e da "Inteligência Artificial", como ser "a" referência no conhecimento sobre o mundo.

Outra interessante premissa epistêmica é a descrição do senso comum como universal, acessível, e não-problemático: "o tipo de conhecimento básico que podemos assumir que agentes humanos possuem"(Taylor et al., 2007). Os proponentes de Cyc sugerem que o "mundo" pode ser acessado e entendido de forma consensual e determinada, independente de quem está engajado no processo de entender. Da forma como é apresentado, esse conhecimento também é considerado enunciável, isto é, pode ser dito e representado de forma "direta" e simplificada. Ao mesmo tempo, o fato de que um grupo restrito tenha a posse e o acesso a esse conhecimento não parece algo questionável, assim como o fato de que alguns sistemas não teriam o apoio desse cabedal.

Lenat expressa dessa forma em quê consiste esse conhecimento geral e universalmente difundido que deve preencher a base de dados de Cyc:

> […] vamos dizer ao computador todos os tipos de coisas que você sabe sobre carros, cores, a Torre Eiffel, a altura dos edifícios, e filmes, e assim por diante (Google TechTalks, 2006).

É interessante observarmos que a expectativa de "acumular" um senso comum universal não surpreende demasiadamente a alguns cientistas. Entretanto, a aparente consensualidade mesmo sobre esses itens tão simples e conhecidos de nosso mundo pode ser colocada em xeque a partir de alguns questionamentos sensíveis – e que nem por isso deixam de ser sensatos. Essa lista deixa entrever o universo de conhecimento de quem formulou a lista, um universo em que há edifícios e no qual é sabida sua altura, em que a Torre Eiffel participa como objeto arquitetônico amplamente conhecido. Nesse universo, é natural que carros sejam importantes e pessoas saibam muito sobre esse tema (Interrante, 1983), e as cores de objetos são indiscutíveis (Goodwin, 1995).

O problema que se coloca é como decidir quais fatos e quais regras são indiscutíveis – um carro não tem o mesmo significado em todas as culturas tal como na cultura americana, e, em que pese a globalização, tampouco *são* os mesmos em todos os lugares (no Brasil, por exemplo, os motores de carros saem de fábrica aptos a serem abastecidos com álcool ou gasolina indiferentemente). Dada a naturalidade com a qual se assume um conhecimento incontrovertido sobre estes objetos, vale perguntar: como e por que um conhecimento como esse pode ser considerado universal, excluindo uma série de pessoas que poderiam não compartilhar deste saber? Qual é a consequência para uma pessoa que falhar em compartilhar desse conhecimento (dito) universal, ou seja, qual o efeito político desse conhecimento técnico? Tudo se passaria como se esse conhecimento definido como sendo de senso comum também fosse independente dos domínios dos fazeres e das linguagens que os constituem, ou seja, auto-evidentes sem a participação do observador.

Acompanhando esta demarcação do que seria ou não legítimo, verifica-se também a legitimação de *um determinado* corpus como válido e relevante. Em outras palavras, quando marcamos a Torre Eiffel como simplesmente "coisas que pessoas sabem", estamos estabelecendo de forma sutil ao mesmo tempo duas relações sobre as quais se deve refletir com cuidado. Em primeiro lugar, estabelecemos a necessidade de conhecer tais objetos culturais para pertencer à categoria de "pessoas que sabem" e para ser capaz de "sobreviver no mundo real", e em segundo lugar, estabelecemos que objetos culturais, conhecidos por diferentes grupos e que *não estão* no corpus, não são importantes para agir competentemente no mundo ou para ser uma pessoa que *sabe*. A demarcação do conteúdo não é explicitada no projeto de Cyc. No entanto, contrastando com a simplicidade alegada para esse conteúdo – aquilo que até "Og, o homem das cavernas" sabia (Google TechTalks, 2006) – o projeto é um complexo empreendimento científico-tecnológico. Para dar conta do desafio a que se propôs, elaborou uma linguagem lógica expressiva que tornasse factíveis as tarefas de raciocínio e inferência utilizadas para manipular as declarações e relações que compõe sua base de conhecimento. Lenat descreve que o problema da coexistência de afirmações não consistentes entre si, ou mesmo contraditórias, não demorou a surgir no decorrer do projeto (Panton et al., 2006). A solução encontrada para dar conta do problema que essa situação representava para Cyc foi a criação de micro-teorias. Micro-teorias em Cyc são contextos de raciocínio, dentro dos quais afirmações devem ser consistentes. Micro-teorias são organizadas em uma estrutura de árvore (ou

mais especificamente, em um grafo direto); aquelas de "alto nível" são as mais genéricas, as quais são progressivamente subdivididas em contextos mais específicos (TAYLOR et al., 2007). A divisão estanque entre conjuntos de afirmações aparentemente contraditórias, no entanto, é uma solução que satisfaz o problema de lógica, de representação em termos computacionais. Não há evidência de que esse tipo de compartimentação da realidade seja a forma como tratamos o nosso estar no mundo mediado pelo conhecimento de senso comum (conforme discutido, a respeito de Cyc, por Adam).

Adicionar conhecimento e manipulá-lo, portanto, longe de parecer-se com o uso cotidiano do senso comum, é uma atividade especializada que demanda um saber específico e não universalmente acessível, que é o da lógica matemática e sua aplicação à computação. Esse fato ajuda a elucidar quem de fato adiciona conhecimento e o manipula, em Cyc. Há várias indicações de que o conhecimento que abarrota Cyc tem sua origem nos saberes de especialistas, e é selecionado e estruturado em função dos valores e particulares perspectivas desses grupos de especialistas. Naphade (NAPHADE et al., 2006) descreve um projeto, com a participação da equipe de Cyc, que propõe estabelecer uma ontologia para classificar conteúdo de multimídia, incluindo as categorias adequadas para descrever esse conteúdo. Segundo o artigo, foram reunidos nesse esforço em particular "especialistas de diversas comunidades para criar uma taxonomia de 1000 conceitos para descrever vídeo-noticiários". O projeto relaciona-se com Cyc de diversas maneiras, incluindo a validação dessa ontologia comparando-a com relação de conceitos já presentes em Cyc. A participação de especialistas é essencial para assegurar que as categorias descrevam de forma adequada e relevante conteúdo de forma que possa ser acessado e classificado de diversas maneiras, incluindo métodos automatizados. No entanto, o resultado não é um sistema de categorias utilizado cotidianamente, próximo ao senso comum (seja de qual grupo for) – mas sim um sistema complexo, altamente estruturado, adequado para as necessidades de especialistas e particularmente para o tratamento computacional dessa informação. A perspectiva do pesquisador enquanto observador não é sublimável em uma perspectiva externa, "de lugar nenhum"; por esse motivo, e em especial quando se busca apagá-la, essa posição de observador ressurge inscrita no conhecimento construído.

De maneira semelhante, são frequentes as menções à dificuldade de se lidar com conhecimento não-especialista (i.e., senso comum) e com as pessoas que o detêm. Desde a

observação de que "especialistas de área geralmente não distinguiam elementos genéricos de instanciados sem serem instruídos" (WITBROCK et al., 2005), até o relato de que

> historicamente, todas as afirmações foram incluídas ao Cyc por ontologistas, engenheiros de conhecimento humano que estão familiarizados com a estrutura de Mt [micro-teorias] e capazes de determinar o posicionamento adequado em Mt (TAYLOR et al., 2007)

ou a descrição do processo de representação como "doloroso", há muitas indicações de que o senso comum não é facilmente "capturável". Para que o conhecimento em Cyc esteja formulado da maneira apropriada, é necessária a intervenção de especialistas, que afinal constituem a população a partir da qual este conhecimento é obtido. O conhecimento em questão é válido, mas dificilmente pode ser considerado "senso comum" e "universal".

Lenat propõe outras maneiras para constituir o conhecimento em Cyc, em parte para diminuir o que é encarado como um gargalo no projeto – a lenta e dolorosa (PANTON et al., 2006; WITBROCK et al., 2005) tarefa de entrada de dados por especialistas. Uma destas é um portal na internet[7] , aberto à participação de voluntários, para coletar "fatos". O portal formula uma série de afirmações a respeito dos conceitos presentes em sua base. Essas afirmações são apresentadas, em um formato de "jogo", aos voluntários – "noviços", isto é, não-ontologistas ou especialistas. Os participantes então devem avaliar e validar – dizer se são válidas ou não – essas afirmações.

O portal, conforme proposto, é uma maneira interessante de abrir o projeto à participação de não-especialistas, e de contar com o precioso (e difícil) conhecimento comum e cotidiano. Há algumas barreiras, entretanto, à participação em Cyc, e, embora não sejam imediatamente visíveis nem tampouco sejam abertamente discutidas pelos proponentes, pode ser instrutivo deter-se um momento sobre elas. A primeira é o fato de o portal ser escrito em inglês, o que coloca uma exigência nada universal – mesmo que não seja extraordinariamente restritiva. A segunda é o próprio fato de ser um portal internet, fazendo do acesso a um computador e à rede uma necessidade para a participação. Por fim, deve ser levado em conta o contexto e a natureza do projeto: é um projeto computacional desenvolvido dentro da disciplina muito especializada de Inteligência

7 Chamado de *FACTory,* em http://opencyc.com

Artificial e divulgada basicamente dentro dessa comunidade. Ser parte dessa comunidade não é um requisito formal para o engajamento, mas chegar ao portal e participar no desafio proposto é mais acessível para quem partilha dos conhecimentos e motivações da comunidade.

A iniciativa do portal na internet, além da oportunidade de uma pequena participação nesse fascinante projeto, permite ao interessado um vislumbre mais direto sobre a visão de mundo construída nessa base de conhecimento. Ao participar da enquete, fui inquirido sobre o íon cromato - "todo o íon hidrogênio cromato tem exatamente um átomo de hidrogênio". Minha formação como engenheiro tornava esse tema ao menos um pouco familiar – embora não tivesse certeza sobre a afirmação no momento. Após alguns tempo de pesquisa pude resolver a dúvida e responder "sim". A natureza científica e estruturada do conhecimento em questão é visível – trata-se de um tipo específico de verdade corriqueiramente manipulada por um grupo também específico. Uma das questões a seguir, em contraponto, coloca em evidência o problema do senso comum: "Todo pudim inclui algum sal de cozinha". Um exame cuidadoso da questão traz à tona o seu caráter não determinável, ou seja, a dificuldade que há em dar uma resposta definitiva e universalmente válida a ela. Saberes da cozinha são enunciados e praticados de maneira diferente daqueles da ciência, e a adição de (um bocadinho) sal a receitas doces é considerada necessária em algumas tradições familiares e em outras não. A adição de sal pode não constar da receita escrita, mesmo que recomendada pela tradição. Nesse caso, uma resposta "sim", tanto como uma resposta "não", seriam respostas válidas, mas extremamente parciais, que ganhariam validade apenas se explicitadas dentro de um corpo referente a uma tradição culinária. A inclusão ou a interdição de ingredientes na alimentação remetem a ressonâncias em outros âmbitos da vida social, não sendo puramente utilitários (Douglas, 1972) e isso também é válido para o sal em nossa sociedade contemporânea (Arnaiz, 2001).

Também presentes, na rodada de questões selecionadas pelo sistema, estavam questões sobre cultura: "Christopher Petiett aparece em 'Boys'", e "'The Clown' foi originalmente escrito em inglês". Dado o fato de que a linguagem que permeia Cyc ser o inglês, não é estranho que estejam em pauta obras culturais em inglês. De uma perspectiva um pouco mais distante, no entanto, é visível, pela maneira sucinta em que foram escritas as declarações, que não se espera que haja dúvidas ou ambiguidades a

respeito do contexto das obras ou seus autores. Esse não parece ser o caso, de qualquer forma, mesmo entre a comunidade de validadores de Cyc. Após respondermos "não sei" à primeira questão, o sistema admitiu, de maneira algo surpreendente: "Ok, acho que isso é mesmo muito obscuro".

Um conjunto de conceitos de Cyc relacionados a "trabalho" constitui um caso adequado para investigar a marca do conhecedor em uma base de conhecimento que reivindica um conteúdo "universal". O conceito "trabalho" é descrito sumariamente como um exercício individual:

Atividades que requerem um certo aporte de esforço físico ou mental, e que não são realizadas puramente por recreação, mas como parte de uma ocupação (para ganhar a vida), ou para contribuir para algum objetivo, ou por causa de alguma outra obrigação ou necessidade. Exemplos incluem um estudante fazendo trabalho de casa para a escola, um proprietário aparando seu gramado, um professor dando aulas, etc[8].

Nada há aqui que indique as dimensões sociais do trabalho, como um fenômeno de construção de realidades e de riquezas percebido em uma escala além do sujeito individual atômico e autônomo. Trabalho, na definição aqui encontrada, é uma atividade não problemática e bem delimitada. A definição explicita, de maneira abstrata mas que denota uma preocupação com abrangência, uma série de motivações para o trabalho. No entanto, os conceitos relacionados que podem ser encontrados em uma busca nessa base vão compondo um quadro de compreensão bastante específica, nada "universal", do fenômeno trabalho: "semana de trabalho" ("cinco instâncias sucessivas de DiaCalendário") e "dias úteis" ("dias em que trabalhadores de escritório e operários normalmente vão ao trabalho e em que estudantes [...] vão à escola"), "plano de trabalho" ("coleção de documentos que contêm informação sobre quando cada empregado deve trabalhar"), "posto de trabalho" – "inclui local da caixa registradora, balcões de expedição, estações de máquinas operatrizes, balcão de trabalho de cozinha (com todos os equipamentos)". Esse conjunto de conceitos, formulados dessa maneira, aponta para uma descrição adequada, mesmo que purificada de elementos controversos, do mundo do trabalho, mas que é também muito específica. Vemos aqui um universo de trabalho organizado e institucional, com empregados (e empregadores), regido por documentos, organizado e distribuído no

8 http://sw.opencyc.org/concept/Mx4rvVjhIpwpEbGdrcN5Y29ycA

tempo da semana e no espaço do local próprio do trabalho. Para tornar ainda mais visível o ponto de vista nada universal a partir do qual esse trabalho é visto e descrito, é instrutivo observar a definição de "visto de trabalho": "Um visto de trabalho nos Estados Unidos, permissão do governo dos Estados Unidos para um não-cidadão trabalhar no país". O conceito *visto de trabalho,* aparentemente genérico, torna-se particular, específico na sua definição, e sua vinculação geopolítica torna-se explícita – e informativa, para além do recorte inicial da intenção dos seus formuladores.

Retomando a discussão de Adam sobre o sujeito conhecedor implícito considerado por projetos tais como Cyc, a descrição de mundo encontrada em Cyc, como vemos, pode ser examinada com atenção para dizer muito sobre os sujeitos reais que se dedicaram a essa tarefa. Formalizar o conhecimento sobre o mundo é materializar em declarações pontuais um saber que é difuso, no sentido de que não é formulado ou formalizado antes de colocado em questão. Cada declaração responde a uma pergunta, uma problematização que por si já é uma indicação de relevância do ponto específico problematizado para o sujeito que realiza esse trabalho. Os conceitos em Cyc aqui apresentados mostram, como aponta Adam, que o saber está intimamente implicado com o sujeito sabedor, em um regime de verdades no qual esse saber se constituiu. Ambos podem adquirir a aparência de separação apenas quando, por esforço explícito ou por condição implícita, a alteridade e a multiplicidade de outras comunidades de saber são invisibilizadas ou desconsideradas. Por outro lado, quando a invisibilidade do saber *outro* é a condição estabelecida e o saber legítimo parece universal, é preciso um olhar cuidadoso para fazer aparecer as fissuras a partir das quais situar e localizar esse conhecimento e os sujeitos a ele ligados.

## *5.5 A materialização de uma perspectiva específica*

Cyc, como um projeto de sistema computacional, é uma proposta audaciosa de construção de um programa capaz de lidar com um domínio difícil: o do conhecimento cotidiano, o *senso comum*. A consideração atenta do projeto sugere que o senso comum ali é figurado como universal, acessível e consensual. Ao que apresentá-lo dessa forma, seus proponentes apagam formas alternativas de conhecer e agir, privilegiando uma visão de

mundo bastante específica, de um grupo com características particulares que é a comunidade da Inteligência Artificial.

O saber representado e incluído no projeto é construído na expectativa de ser corriqueiro e não-problemático. Essa situação, e o fato de que os sujeitos participantes no empreendimento compartilham um envolvimento com a comunidade que lhe dá suporte, tornam de certa maneira opaca sua filiação, e menos visível a maneira em que é específico. Por esse motivo propomos interrogar esse saber, e tornar observáveis as marcas de sua origem e dos pressupostos que o orientam. Esse questionamento é importante na medida em que as marcas e pressupostos que observamos são compartilhados por uma série de outras realizações tecnológicas computacionais cuja funcionalidade e eficiência também são naturalizadas e tornadas opacas.

Estamos considerando aqui uma linha específica – a de sistemas especialistas, ou sistemas baseados em conhecimento – dentro da área científico-tecnológica mais ampla da Inteligência Artificial (que inclui a, mas não se limita à, tradicional IA simbólico-lógica). É uma abordagem estabelecida já há algum tempo, desde a década de 1970 (ver acima). No entanto, permanece relevante na medida em que continua sendo, atualmente, a escolha adequada para basear o projeto de inúmeros sistemas de informação computacionais, tendo recebido nesse meio tempo inúmeros aportes teóricos e tecnológicos. Colocamos aqui sob escrutínio uma realização específica, um artefato tecnológico em particular, um sistema especialista de senso comum. O objetivo foi mostrar como alegações a respeito de generalidade e universalidade de conhecimento e eficácia tecnológica são constituídos, em um complexo jogo que vai desde premissas epistemológicas fortes e não explicitadas até a apresentação do sistema computacional de maneira a ressaltar a obviedade do "universal" e a apagar o "particular" e a vinculação do artefato a realidades peculiares dos sujeitos conhecedores e construtores. Consideramos essa análise pertinente, na medida em que outras situações semelhantes, isto é, de apagamento de multiplicidades epistemológicas ligadas ao poder de construir tecnologias eficazes, podem ser identificadas tanto dentro das diversas linhas da Inteligência Artificial como em outras áreas de realização técnico-científica.

O esforço devotado ao desenvolvimento de Cyc reverbera dentro da área da Computação e da Inteligência Artificial, assim como em outras áreas tais como a Filosofia,

a Antropologia e os Estudos de Ciência e Tecnologia, em função das características peculiares do projeto. São características que ancoram mutuamente questões de epistemologia, tecnologia avançada, atributos humanos privilegiados em nossa cultura tais como inteligência e agência, além de fortes – mesmo que implícitas – ressonâncias políticas e de poder relacionadas a regimes de saber. Essa situação faz da produção tecnológica da Inteligência Artificial um campo de interesse, no qual é possível observar a materialização de formas específicas de compreender o sujeito e o conhecimento, e não apenas realizações de características humanas consideradas genéricas ou "universais".

# 6 O jogo da interpretação entre humanos e agentes artificiais[9]

A ideia de tornar sistemas computacionais agentes de diversos processos sociais, nos quais os participantes são usualmente seres humanos, é uma das motivações para a pesquisa e inovação no campo da Inteligência Artificial. O objetivo é que estes sistemas possam ser agentes em relações características tais como prestação de serviços, aprendizagem, e mesmo em relações de afeto, ao imbuí-los de determinadas características e torná-los, através da programação explícita, intrinsecamente aptos para sociabilidades.

Esta abordagem tem rendido frutos notáveis, tanto no âmbito da simulação e recriação de formas de aprendizagem próprias do mundo social, como no âmbito da relação social ampliada de humanos com máquinas. No entanto, o que temos procurado argumentar é que a ênfase em modelos prévios não deixa espaço para momentos de interação não esperados ou não acomodáveis dentro do modelo. Estes momentos são figurados como falha, ou como problema, como erros cuja responsabilidade última deve ser encontrada em uma falta da máquina ou do modelo, ou em um erro do utilizador. Propomos, como uma possível alternativa, uma perspectiva que reconheça o caráter generativo e criativo destes momentos, reconhecendo a maneira como as práticas de sociabilidade e interação humanas podem apropriar-se destes momentos como um recurso para a interação com o mundo, e não como um obstáculo a uma intenção subjacente, pura e bem formada.

Apresentaremos, neste capítulo, um momento do encontro entre artefatos computacionais da Inteligência Artificial, usuários destes artefatos, e pesquisadores que

9 Uma versão do material que constitui este capítulo foi publicada como (Wild et al., 2010)

desenvolvem estes artefatos. O sistema aqui descrito, o agente artificial que chamávamos coloquialmente de Gato, foi desenvolvido por Gepê, um dos pesquisadores do grupo. Quando cheguei ao grupo, o procedimento experimental que será havia recém sido efetuado, com resultados inconclusivos que deixaram os pesquisadores algo perplexos – afinal, um sistema projetado e construído sob as regras da engenharia não realizou o objetivo ao qual foi proposto. Quando entrei em contato com Gepê e conheci alguns detalhes do sistema e dos resultados, imediatamente interessei-me justamente por este aspecto: a contradição aparente entre o projeto cuidadoso de um sistema e resultados desorientadores quando de seu encontro com pessoas.

Ao longo de meu tempo em campo, envolvi-me com o projeto, em um desenvolvimento do mesmo sistema com alguns requisitos diferentes. Concomitantemente, trabalhamos eu e Gepê para tentar encontrar sentidos na aparente desordem deixada pelas respostas dos usuários. O que emergiu foi a proposta, singela mas significativa como abertura, para uma abordagem diferente das interações de humanos com máquinas, com o intuito de abrir espaço para a diversidade e o não conforme. É por isto que este capítulo fecha o ciclo de questionamentos colocados por esta tese.

Para tanto, interrogamos, com o auxílio de um conjunto de premissas que são propostas como uma alternativa à abordagem usual da IA, um caso experimental, em que investigou-se a atribuição de um julgamento cognitivo social, a identidade, a um agente artificial capaz de migrar entre plataformas diferentes. Proporemos a seguir algumas considerações sobre a relação entre modelos computáveis na inteligência artificial e o espaço aberto da sociabilidade humana.

## *6.1 Novos argumentos interpretativos*

A atribuição de uma identidade estável a um ente, isto é, reconhecê-lo como um ente dentro de uma ontologia que dele dê conta, é um problema filosófico complexo, mas surpreendentemente parece ser um processo transparente no cotidiano das pessoas. De alguma forma, as pessoas dão conta, no dia-a-dia, de reconhecer e conceder continuidade

existencial aos entes que as circundam, sendo capazes de discernir as capacidades e as diferentes agencialidades dos entes com quem travam relações.

Estendendo a discussão que iniciamos no capítulo anterior, é possível observar que há uma pródiga variedade de ontologias de agentes, isto é, de elencos de entes aptos a entrar em relação com os humanos e o tipo de relações esperadas destes entes, que podem ser encontradas em diversas sociedades e grupos culturais. Para além da ontologia, talvez familiar a muitos de nós, em que entram em consideração apenas seres "humanos" autônomos e racionais, há ontologias em que os humanos entram em relações sociais ativas com orixás e outros espíritos (Bastide, 2001); há a complexa intersubjetividade humano-animal (Despret, 2004); ou em que humanos relacionam-se com programas de computador ou outras máquinas computacionais (Castañeda & Suchman, 2005; Weizenbaum, 1966).

A questão do reconhecimento de um outro como sujeito de relação, de intersubjetividade, entre humanos e não-humanos, foi investigada por Vincianne Despret (Despret, 2004), a propósito da questão da separação ou da distinção de um "nós" humano. Ao etnografar criadores de animais, Despret aprendeu com seus sujeitos como a questão da separação não é a mais importante em todas as situações. O que surgiu em sua descrição da atividade destes criadores é a translação da questão da diferença entre o que é humano e o que é animal para outra questão, mais sensível e mais importante no mundo em que estão as pessoas pesquisadas: a diferença entre situações, nas quais há maneiras diferentes de humanos e animais relacionarem-se entre si, e de proporem-se mutuamente subjetividades com as quais relacionar-se. Ao dar oportunidade, para os criadores com os quais trabalhava, de ajudá-la a entender o que relevava naquilo que constituía-se seu universo de relação, Despret chegou a uma conclusão interessante. A questão original de pesquisa viu-se insuficiente, isto é, para os sujeitos da relação, a caracterização externalista dos entes envolvidos não bastava, não descrevia o que sentiam como relacional. Entenderem seu modo de relacionar-se com seus animais (pequenos criadores com relação próxima com seus animais) como diferente de outros modos (grandes fazendas em que os animais são ignorados enquanto entes individuais), isto sim os produzia como os sujeitos desta relação. Produzia também, como sujeitos relacionais, os animais que criavam.

Ao jogar com a criação de sistemas complexos, nos quais consegue ver em ação características que reconhece como próprias suas, o ser humano abre caminhos para relações, com estas máquinas, nas quais vai reconstruir de diversas formas os traços pelos quais reconhece suas relações com humanos. Uma descrição de engajamento afetivo e social é a história de Lucy, um robô que lembra um bebê orangotango, desenvolvido ao longo de vários anos por Steve Grand (GRAND, 2004). A visão de Grand é a de desenvolver a inteligência robótica de Lucy como a de uma criança, ao longo de um período. O projeto é apresentado de maneira peculiar; Grand constrói para Lucy uma vida social, incluindo um diário na internet e fotografias de família. A história que Grand conta retoma figuras e ações retóricas que remetem a categorias de criança, desenvolvimento infantil e relacionamento familiar próprios de seu universo social e sua posição como cientista, homem, branco (CASTAÑEDA & SUCHMAN, 2005) – e remete também a expectativas que informam sobre suas noções de agencialidade: "[my dad is] *out ready to start building ... the very large neural network ... to make me into a complete organism. After that it's up to me*"[10] (grifo nosso).

Dada esta variedade de ontologias existentes e de entes que as povoam, é de se perguntar se bastam características intrínsecas aos entes para que possam ser reconhecidos de maneira estável. Há evidências de que, mais do que simplesmente reconhecer uma agencialidade existente, os seres humanos colocam em ação uma série de métodos pelos quais reconhecem a existência de uma ordenamento do mundo, métodos estes que, quando utilizados, instauram em um movimento recíproco o mundo ordenado. São chamados de etnométodos por Garfinkel (COULON, 1995; GARFINKEL, 1984), por serem são os métodos através dos quais cada grupo culturalmente diferente identifica e recursivamente constrói o cotidiano inteligível. Para Garfinkel as pessoas, no decurso de seu dia-a-dia, utilizam uma forma de "ler" e "compreender" a realidade, chamada por ele de "método de interpretação documentária". Este método funciona com as pessoas isolando e categorizando certos indícios como sendo características, dentro do plano geral de suas percepções, de acordo com uma teoria subjacente que explica estas características. As características percebidas servem então recursivamente como justificativa e confirmação para a teoria.

---

10 http://web.archive.org/web/20040312134720/www.cyberlife-research.com/about/

A importância de considerar perspectivas alternativas, tais como a etnometodológica, para abordar as sociabilidades do artificial, está em problematizar o processo de elencar características "necessárias" ou "suficientes", interessando-se pela forma como estas características são apropriadas pelo utilizador e as maneiras pelas quais utiliza tanto as características modeladas como outras propriedades inesperadas do objeto em seu julgamento de identidade. A ênfase muda, da urgência em completar um catálogo de essencialidades para a curiosidade em observar como o que está à mão é usado como recurso pelo utilizador engajado na interação. Abrir espaço para inquirir a maneira como o sujeito participante constrói a interação é também uma tentativa de reconhecer a assimetria que existe na relação entre o humano e a máquina, sem figurá-la em termos de insuficiência ou de déficit (SUCHMAN, 2007, p. 269). Dentro desta perspectiva, a prioridade, ao invés de recuperar para o artificial uma igualdade com o humano, torna-se realizar dispositivos tecnológicos como lugares de relações cada vez mais criativas com o humano.

## *6.2 Uma construção experimental direcionada para o reconhecimento de um ente*

O ponto de partida para esta análise foi a criação, utilizando técnicas de inteligência artificial, de um agente artificial, animado, capaz de jogar xadrez (Cuba, 2010), como um projeto realizado dentro do grupo de pesquisa português. Este agente convida o usuário a uma partida, comenta em voz alta sobre as jogadas, anima suas expressões faciais como reflexo do que está a acontecer na interação do jogo de xadrez com o usuário, e uma vez terminado o jogo convida o usuário para uma próxima partida, demonstrando através de comentário em voz alta que recorda-se do resultado da partida jogada. Este agente possui adicionalmente uma característica que lhe é própria, que é a de rodar em plataformas diferentes. Roda em um robô – iCat, um robô com feições faciais animáveis (VAN BREEMEN, 2005; VAN BREEMEN et al., 2005) – e em um telefone celular com capacidades gráficas, onde foi dotado de uma representação visual de seu rosto e de um tabuleiro de xadrez. Mais notavelmente, o agente é capaz de migrar de uma plataforma para outra, isto é, respondendo a uma solicitação de um usuário, o agente a rodar em uma dada plataforma suspende a interação em curso, exibe uma animação que sugere que a plataforma em

questão não está mais "animada" por um agente, e passa a animar a plataforma diferente que foi escolhida pelo usuário. A interação passa a partir deste momento a ocorrer na nova plataforma, continuando o jogo que que estava em curso anteriormente.

A experiência foi realizada com o objetivo de conhecer a influência que certas características do agente poderiam ter na percepção, por parte dos usuários, do agente como sendo "o mesmo" em uma plataforma e outra, isto é, da identidade do agente como percebida pelo utilizador. Algumas características foram previamente postuladas como de interesse para investigação: a aparência; a personalidade; a voz; e a memória de interações passadas. Estas características eram moduladas, isto é, programadas com diferenças sistemáticas entre uma plataforma e outra, para a realização do estudo. A aparência foi modulada através da imagem que era apresentada no telefone (um desenho do robô iCat ou um desenho tipo *cartoon* de uma pessoa) em contraste com o robô físico. A personalidade foi, para o presente caso, construída como um conjunto simples de demonstrações de humor (e.g. triste, contente) presente na animação da expressão facial do agente, e como uma série de reações a eventos, evidente nas frases enunciadas pelo agente em situações do jogo tais como tentativas de jogada ilegal ou captura de peças – o objetivo não era obter uma complexa construção, no sentido da ciência psicológica, de personalidade, mas construir condutas que fossem interpretáveis como relacionadas a atitude putativa do agente. Uma modulação da personalidade era "amigável", a outra era "antipática". A voz em um caso e outro era proveniente de um sintetizador de voz e de gravações das falas geradas pelo sintetizador, em língua inglesa; uma das vozes soava com acento americano e a outra soava com acento europeu. Uma característica do tipo "memória" foi implementada com o armazenamento do resultado do último jogo e disponibilizando o resultado, na fala do agente, no momento do início do jogo seguinte.

Durante o procedimento do experimento, o participante era solicitado a assistir a dois vídeos com performances em que uma pessoa jogava xadrez com o agente. O primeiro vídeo consistia de uma pessoa a jogar xadrez com o agente que rodava num robô iCat. O segundo vídeo, do qual havia cinco variantes, também mostrava uma pessoa jogando xadrez, mas seu oponente era o agente rodando em um telefone celular com tela gráfica. No primeiro vídeo, a aparência visível do agente era a do robô, sua voz sintetizada – com as falas em inglês – soava com acento americano, e a personalidade era programada para parecer "amigável". Participantes então eram atribuídos aleatoriamente a um dentre cinco

grupos experimentais. No segundo vídeo apresentado as características do agente variavam, sendo moduladas consoante o grupo experimental. Foi realizado um grupo experimental para cada característica que pretendíamos avaliar, e um adicional para propósitos de controle da experiência. Cada grupo assistia a um vídeo em que o agente no telefone apresentava *uma* característica em comum com o agente no primeiro vídeo, que era a característica que pretendíamos avaliar nessa condição, e as restantes características eram moduladas para tornarem-se *diferentes*. No caso da condição de controle todas as características que modulamos na segunda apresentação do agente são diferentes.

Para realizar a experiência utilizamos um questionário online. Este questionário começava por fazer uma breve introdução, referindo-se sucintamente a uma avaliação de personagens que jogam xadrez, não mencionando explicitamente o que se procurava avaliar em concreto. Em seguida o participante preenchia dados indicando o seu gênero e faixa etária. Após isto lhe era pedido para assistir aos dois vídeos – o primeiro vídeo, com o iCat, igual para todos os participantes, o segundo, um vídeo em que o agente estava no telefone celular, escolhido entre os cinco disponíveis, de acordo com o grupo atribuído ao participante.

Após a observação dos vídeos era pedido aos participantes que indicassem numa escala de 7 pontos (do tipo Likert) o quanto achavam que a personagem apresentada nos dois vídeos era a mesma. Em duas questões de resposta aberta, era-lhes solicitado que indicassem livremente em que aspectos haviam achado as personagens respectivamente semelhantes e diferentes. Em seguida, os respondentes prosseguiam para a página final do questionário, em que eram solicitados a classificar uma série de itens expressos, também numa escala de 7 pontos, como sendo semelhantes ou diferentes entre as duas personagens. Estes itens incluíam as características que nós pretendíamos avaliar, assim como outros aspectos da interação (expressões faciais, frases utilizadas, estado do jogo e humor).

O conjunto de participantes consistiu de um total de 71 respondentes. A chamada para o questionário foi realizada pela distribuição de uma mensagem dentro de uma universidade técnica portuguesa, e foi colocada uma mensagem em um fórum internet de jogos de computador. Há uma questão interessante em relação à distribuição de gênero: a predominância é acentuada de respondentes "masculino" (82%) em relação a "feminino" (18%). A proporção de alunos homens na universidade é de fato muito maior do que a de

alunas, mas a proporção dentro do universo de participantes do fórum não é conhecida. De qualquer forma, a proporção de respondentes obtida é uma informação interessante na medida em que se constitui uma informação, mesmo que imprecisa e indireta, sobre distribuição de gênero no universo em questão: um universo da técnica, da computação e da informática.

## *6.3 Respostas recebidas: semelhanças, diferenças*

Como mencionado acima, os pesquisadores envolvidos tendiam a esperar que os usuários fossem "reconhecer" o agente em ambas as plataformas como sendo "o mesmo", em diferentes graus, dependendo das características geradas em comum para os dois vídeos. Por sua vez, isto determinaria, para cada característica, um grau comparativo de importância para o reconhecimento de uma "identidade", no sentido de "ser o mesmo", para o agente em diferentes dispositivos. Os pesquisadores tendiam, da mesma forma, a esperar que usuários não fossem reconhecer o agente como "o mesmo" nos casos em que a similaridade – da forma como idealizada no experimento – fosse menos óbvia ou não fosse presente.

Não foi este exatamente o caso. Comparando os quadros 1 e 2, não há uma situação em que claramente os usuários reconhecessem ou deixassem de reconhecer o agente como "o mesmo" (nestes quadros, as respostas, originalmente expressas na escala numérica, estão agrupadas e organizadas em 3 categorias, "não é o mesmo", "neutro" e "é o mesmo"). O valor numérico respondido para cada uma das características, em cada grupo, tampouco correlaciona-se significativamente com o julgamento de identidade – para detalhes estatísticos, ver (Cuba, 2010). Em outras palavras, o item de identificação programado para ser "semelhante" ou "equivalente", entre as versões robótica e telefone do agente foi considerado muito diferente por vários usuários. O quê aconteceu? Os participantes não entenderam o jogo proposto? A situação, inesperada, causou certo desconcerto em nós, pesquisadores.

| *É o mesmo?* | *Semelhanças* | *Diferenças* |
|---|---|---|
| *Não é.* | A tonalidade da voz, embora o discurso fosse mais agressivo. | na atitude, expressão facial que é fortemente influenciada pela boca. |
| *Não é.* | Ambos reagiram ao usuário. Ambos tinham a mesma aparência. [They both reacted to the user. They both had the same appearance.] | O personagem no robô era legal enquanto que o personagem no celular era desagradável. [The character in robot was nice while the character in mobile phone was unpleasant.] |
| *Não é.* | Visualmente eles obviamente parecem o mesmo [Visually they obviously look the same] | Sua personalidade era muito diferente. No primeiro vídeo ele é bondoso. No segundo ele é arrogante e insuportável. [His personality was very different. In the first video he's kind. In the second video he's arrogant and obnoxious.] |
| *Não é.* | Eles eram similares em aparência e estilo de jogo [they were similar in appearence and playing style.] | O personagem do telefone era muito rude na maneira de abordar o jogador, o robô era muito polido [the phone character was very harsh in the way he addressed the player, the robot was very polite.] |
| *Neutro.* | Humor [Humor] | Voz, Tristeza/Alegria [Voice, Sadness/Hapiness] |
| *Neutro.* | Aparência, Voz, Expressões da Face [Appearance, Voice, Face Expressions] | O atraso entre cada movimento do xadrez! Os movimentos da cabeça do gato sugerem outro personagem [The delay between each chess move! The movements of the cat's head suggest other character.] |
| *Neutro.* | Forma, Cor, Detecção de Jogadas Ilegais [Shape, Color, Illegal Move's Detection] | Expressão, Técnica de Jogo [Expression, Game Technique] |
| *É.* | Ambos tinham a mesma face. [Both had the same face.] | O robô parecia mais engajado no jogo. Os olhos vasculhando o tabuleiro faziam-no parecer mais envolvido no processo. A versão no celular parecia ser mais desconectada do contexto do jogo. A versão no celular era também mais irritante/insolente/arrogante. [The robot seemed more engaged in the game. The eyes scanning the board made him seem more involved in the process. The mobile phone version seemed to be more disconnected from the game context. The mobile phone version was also more annoying/cocky/arrogant.] |
| *É.* | As expressões faciais das duas personagens são bastante semelhantes. | A personagem do telemóvel parece um bocado mais agressiva. |
| *É.* | Os traços físicos são similares [The physical traits are similar] | O “ao vivo” parece mais amigável, embora menos interessante [The live one seems friendlier, though less interesting] |
| *É.* | Quase idênticos porque provavelmente ambos tinham a mesma IA [Almost identical because They both probably had the same AI] | O outro tem um corpo físico, enquanto que o outro é pixels. [the other has a physical body, whilst the other is pixels.] |

*Quadro 1: Grupo "mesma aparência", respondentes que preencheram tanto "Semelhanças" como "Diferenças"*

| *É o mesmo?* | *Semelhanças* | *Differences* |
|---|---|---|
| *Não é.* | a expressao facial é identica no entanto o facto do robot ser 3d, torna o "boneco" mais simpatico. | O boneco no telemovel, parece mais arrogante, e que esta a fazer um favor ao jogar. |
| *Não é.* | Plataforma tecnológica [technology platform] | Voz, face, cores, aspectos físicos (2D vs. 3D), velocidade em respostas, etc... [voice, face, colors, physical aspects (2D vs 3D), velocity in answers, etc..] |
| *Não é.* | Apenas na última frase: “Não se preocupe, tudo que você precisa é prática.” [Only in the last phrase: "Don't worry, all you need is practice."] | O primeiro personagem levou algum tempo antes de fazer sua próxima jogada, o que o faz parecer mais humano. O primeiro personagem também parece um tanto mais expressivo que o segundo. O primeiro personagem parece jogar melhor que o segundo, as jogadas que faz não parecem tão “kamikaze”. [The first character took some time before making his next move, which makes it look more human.<br>The first character also looks a lot more expressive than the second.<br>The first character seems to play better than the second, the moves it makes don't look so "kamikaze".] |
| *Não é.* | Suas bocas e olhos, similares, e uma voz robótica. [They similar mouth and eyes, and a robotic voice.] | O primeiro era um gato e tinha um corpo (o que lhe dá presença). O segundo era uma cabeça humana, e mais feia. [The first was a cat and had a body (which give him presence). The second was a human head, and uglier.] |
| *Não é.* | Frase similares; algumas expressões [Similar phrases; Some expressions] | Frases diferentes para xeque, e outras situações. O personagem no telefone era menos emotivo. [Different phrase for check, and other situations. The phone character was less emotional] |
| *Não é.* | Ambos personagens apresentam algum tipo de mímica [both characters displays some kind of mimics] | Ambos apresentam mímica mas as que o robô apresenta são muito melhores, ele comporta-se mais como um rival [they both display mimics but the one that the robots display are far more better, it behaves more like a rival] |
| *Não é.* | Bem, ambos sabem jogar xadrez, e todas as regras, assim parece. [well the both know how to play chess, and all the rules, so it seems.] | O personagem apresentado no telefone era mais solícito, porque ele deixava seu oponente desfazer sua última jogada, dando a ele uma nova chance. Também este personagem era apresentado com maneiras mais polidas. [The caracter presented in the phone was more helpfull, because he allow his opponent to undo his last move, providing him a new change. Also this caracter was presented by more polite manners.] |
| *Não é.* | A mecânica e aparência das expressões faciais [the mechanics and looks of the facial expressions] | O segundo personagem é menos interativo [the second character is less interactive] |
| *É.* | As expressoes faciais, alegria, tristeza, pensativo, entre outras, são o que mais aproximam ambas as personagens. | No caso do telemóvel, a personagem é muito mais explicativa, mais faladora e, de certa forma, isso torna-se mais atractivo. |
| *É.* | Em ambos os casos eles agem como | São diferentes na aparência, o gato é mais bonito. |

| *É o mesmo?* | *Semelhanças* | *Differences* |
|---|---|---|
| | amigos/parceiros, alertando sobre as boas e as más jogadas. | |
| *É.* | A mesma maneira com que jogam [The same way they play] | Um é físico, o outro virtual... [One is phisical, the other is virtual...] |
| *É.* | Eles têm as mesmas sentenças, seu sotaque é muito similar, mesmo suas expressões faciais são muito próximas. [They have the same sentences, their accent is very similar, even their facial expressions are very close.] | O personagem do celular não tem os ruídos irritantes que o real tem. [The mobile phone character does not have the annoying noises the real one has.] |

*Quadro 2: Grupo "mesma personalidade", respondentes que preencheram tanto "Semelhanças" como "Diferenças"*

Uma das explicações possíveis seria o fato de que a atividade aconteceu online, e que a experiência com o agente era mediada por vídeo, o que poderia dificultar a correta avaliação, por parte do respondente, da performance das instâncias do agente. Esta é uma explicação plausível, já que a vivência de um evento é qualitativamente diferente da vivência de um vídeo (mesmo que retratando um evento do mesmo tipo, ver por exemplo a discussão metodológica colocada por Flick, 2004, a respeito de pesquisa utilizando gravações visuais). No entanto, não é uma explicação completamente satisfatória, uma vez que o público participante tem como característica a utilização cotidiana do computador, do vídeo e da sociabilidade remota enquanto maneiras de vivência. Vivência suportada na tela do computador não é equivalente a vivência no mundo "real", mas é uma vivência que existe e é exercida com estatuto próprio e rotineiro por um conjunto de pessoas, incluindo o universo do qual compartilham os participantes do estudo. O fato de o estudo ser online e ter sido respondido já aponta para uma certa afiliação do participante a esta forma de vivência. É importante destacar, também, que organizar e atribuir estatuto de pessoa a uma entidade pode não ser um processo aritmeticamente aditivo, pelo que a questão de formular o experimento baseado em características discretas poderia ser sujeita a uma reelaboração.

O ponto central deste capítulo, contudo – e em acordo com a linha desta tese – não é analisar resultados como *insuficiência* da parte de pessoas ou entes artificiais, e sim concentrar-se no trabalho interpretativo realizado por usuários, no encontro com os agentes artificiais, em sua tarefa de responder a indagações a eles apresentadas.

Como veremos, as respostas dos participantes fornecem uma chave interpretativa a partir da qual compreender o processo de julgamento de identidade para o agente a elas apresentado. O resultado final – "parece"/"não parece" – torna-se a conclusão de um processo que pode ser melhor entendido através de um olhar atento para os depoimentos dos participantes. O raciocínio coletivo traz indicações importantes de como sujeitos socializados dentro de um universo de representações e referências virtuais estabelecem e colocam em ação critérios instrumentais de julgamento próprios deste universo. Iremos nos concentrar nas respostas para "mesma aparência" e "mesma personalidade"; os outros grupos seguem um padrão semelhante. As respostas provenientes dos questionários são citadas verbatim, sem tentativas de correção ortográfica ou outras intervenções, preservando a maneira como foram escritas.

## *6.4 Julgando por aparências, buscando diferenças*

No momento de projetar o estudo, os vídeos em que a **aparência** era modulada de maneira similar pareciam ser o par mais óbvio, o mais facilmente visível. Por este motivo, tendia-se a esperar que os participantes neste grupo respondessem em maior número indicando neste questionário o agente como "o mesmo", dada a forma como foi modulada a semelhança.

As aparências... enganam, pelo visto. Escolher "sim" ou "não", "parece" ou "não parece", "é o mesmo" ou "não é", mostrou-se difícil para os julgadores que participaram neste questionário. Neste grupo, as respostas ao item "Aparência Visual" tiveram em valores altos, próximos ao lado da escala "exatamente o mesmo". O julgamento "é o mesmo ou não é", no entanto, foi disperso, com alguns participantes seguros de que o agente era o mesmo e outros, igualmente seguros de que não era o mesmo. A aparência visual semelhante não facilitou decidir em favor da identidade.

As justificativas expostas para os julgamentos efetuados, no entanto, são ricas em indicações para as formas como o recursos disponíveis foram apropriados e interpretados. Quando nos referirmos a respondentes que "identificaram" ou que "consideraram o mesmo" o agente no robô e no telefone, estaremos considerando respondentes que

marcaram 5, 6, ou 7 na escala em que 7 significava "concordo fortemente", quando perguntados se o agente "era o mesmo". Quando nos referirmos aos que "não identificaram" ou que "não consideraram o mesmo" estaremos considerando os que marcaram 1, 2, ou 3 nesta escala. Quando nos referirmos aos que responderam "no meio da escala" consideraremos os que marcaram 4.

A semelhança visual entre o robô e sua representação na tela do telefone foi apontada explicitamente, tanto por respondentes que julgaram ambos como o mesmo, como por respondentes que julgaram ambos não idênticos. No primeiro caso estão comentários como "*Os traços físicos são similares*" e "*Ambos tinham a mesma face*"; no segundo caso estão os comentários "*Visualmente eles obviamente parecem o mesmo*" e "*Ambos têm a mesma aparência*". Ao responder sobre diferenças, muitos dos comentários eram sobre traços de "personalidade" (ou "atitude"). A maior parte dos comentários sobre estes traços indicava diferenças percebidas; por exemplo, os respondentes acima consideraram estes traços como diferença: "*O 'ao vivo' parece mais amigável*", e "*Sua personalidade era muito diferente. No primeiro vídeo ele é bondoso. No segundo vídeo ele é arrogante*". "Expressões faciais" foram relatadas como traço significativo, e é importante notar que esta avaliação recorre a considerações tanto de aparência como de personalidade. Por exemplo, um participante respondeu "*expressões faciais das duas personagens são bastante semelhantes*" e contrastou esta observação com "a *personagem no telemóvel parece um bocado mais agressiva*". Na mesma linha, outro participante comentou como diferença a "*atitude, expressão facial que é fortemente influenciada pela boca*" – é interessante lembrar que "atitude", em seu significado literal, significa "postura corporal". Outra resposta que conecta traços visuais com uma percepção do tipo psicológico é "*Os olhos vasculhando o tabuleiro faziam-no (ao robô) parecer mais envolvido no processo*".

O questionário em que a **personalidade** do agente era modulada de maneira similar no robô e no celular (e as outras características eram moduladas de maneira diferente) trouxe uma situação interessante, que constitui-se em um binário em relação aos resultados questionário "aparência". Da mesma forma como no questionário **aparência**, o resultado final do julgamento dividiu-se entre considerar o agente como "o mesmo" e como "não o mesmo". No entanto, estando ausente a semelhança mais visível, que era aquela entre o robô iCat e a representação como um desenho na tela, observamos que os participantes procuraram encontrar sentido através do exame e da interpretação

cuidadosa do que estava disponível tanto da aparência visível como de impressões da personalidade.

O resultado foi que a personalidade foi comentada algumas vezes como "semelhante", e algumas vezes como "diferente". Os participantes colocaram critérios em ação: assistiram aos vídeos, organizaram o que viam como características das quais dar conta enquanto "aparência", "personalidade" ou outra categoria, e então pesquisaram estas características para poder chegar a uma decisão em termos de "semelhante" ou "diferente". As decisões, portanto, foram obtidas através deste processo de julgamento, não de maneira imediata – de fato, exercendo um certo esforço.

Como a representação pictórica neste caso é uma cabeça humana, do sexo masculino, em estilo *cartoon*, a similaridade visual com o robô não é óbvia. Mesmo assim, semelhança visual foi expressa, como função de expressão facial ("*a expressão facial é idêntica*", ou "*a mecânica e aparência das expressões faciais*") ou de partes da face significantemente relacionadas ("*suas bocas e olhos, similares*"). Ao apontar diferenças visuais, distinções sobre a natureza dos dois entes foram trazidas: "*aspectos físicos (2D vs. 3D)";* ou *"O primeiro era um gato e tinha um corpo (o que lhe dava presença)*"; ou ainda "*Um é físico, o outro é virtual*".

Embora neste grupo os agentes fossem programados com a intenção de transmitir a ideia de "mesma personalidade" (recordando que este fato não foi anunciado aos participantes), as respostas não necessariamente concordaram com isso. Foi comentado, por exemplo, que "*o boneco no figura no telemóvel, parece mais arrogante*", que "*o personagem no telefone era menos emotivo"*, ou que *"o personagem apresentado no telefone era mais solícito ... [e também] com maneiras mais polidas*". Uma pessoa, contudo, aproximou os agentes: "*em ambos os casos eles agem como amigos/parceiros*".

Neste grupo, as respostas não constituíram um contraste claro e organizado como no anterior. Um número de combinações entre características percebidas foi, em lugar disso, usado para investigar a semelhança dos agentes, dando azo a um arranjo mais amplo de comparações, mostrado nas respostas. Algumas respostas incluíram o mesmo tipo de caraterísticas nos dois lados da comparação; por exemplo, um respondente colocou traços visuais para ambos, sendo a similaridade dada como "*Suas bocas e olhos, similares*" (resposta já citada anteriormente) e a diferença como "*O primeiro era um gato e tinha um*

*corpo... O segundo era uma cabeça humana, e mais feia*". Outro participante destacou a expressão corporal: "*ambos personagens apresentam algum tipo de mímica*", contudo "*a que o robô apresenta são muito melhores*". Falas também surgiram como traço perceptível – para um participante, eram "*frases similares*" em contraste com "*frases diferentes para xeque, e outras situações*". Por outro lado, comparar os agentes usando características diferentes também foi usado como estratégia. Uma das respostas trouxe "*a expressão facial é idêntica*", enquanto que "*o facto de que o robô ser 3d, torna o 'boneco' mais simpático. O boneco no telemovel, parece mais arrogante*". Outro respondente escreveu que "*em ambos os casos eles agem como amigos/parceiros*" e que "*são diferentes em aparência, o gato é mais bonito*". Também houve uma resposta comparando a semelhança de "*expressões faciais, alegria, tristeza, pensativo, entre outras*", e a diferença de "*no caso do celular, a personagem é muito mais explicativa, mais faladora*".

Por fim, o estilo e estratégia de jogo também foram usadas como recurso. Participantes questionaram a sequência de jogadas, procurando a distintividade que pudesse ser usada para realizar o julgamento. Dois comentários sobre o jogo são especialmente interessantes. Um participante escreveu "*a mesma maneira com que jogam*"; outro, pelo contrário, enfatizou as diferentes marcas que poderiam ser lidas no estilo de jogar: o robô "*levou algum tempo antes de fazer sua próxima jogada, o que o faz parecer mais humano... parece jogar melhor que o segundo, as jogadas que faz não parecem tão 'kamikaze'*".

## *6.5 Observar, interpretar, justificar*

Os participantes do estudo aqui descrito foram instados a tomar uma decisão: se ambos os agentes artificiais apresentados eram, ou não, "os mesmos". O resultado deste julgamento não foi homogêneo; não foi encontrada uma situação que claramente definisse que pessoas iriam responder "é o mesmo" ou o contrário. Isto não significa, no entanto, que respondentes não fossem capazes de avaliar e interpretar as circunstâncias da apresentação; como o mostram as respostas escritas, as cenas apresentadas foram consideradas e racionalmente avaliadas, resultando nos traços que foram destacados como

relevantes para definir o julgamento. Essa consideração ocorreu em um espaço interpretativo em aberto, não ainda definido para quem o encontrava – traços que deveriam ser observados, a organização cognitiva que os tornava significantes, e as medidas pelas quais avaliar o "diferente" e o "mesmo" entre as duas cenas, tudo isso foi montado e mobilizado pelos participantes para que pudessem chegar a uma decisão. Isto torna compreensível o porquê dos participantes haverem elencado um conjunto de marcas e traços diversificado, alguns dos quais, como a aparência e a atitude da personalidade, haviam sido antecipados pelos pesquisadores (e intencionalmente programados nos agentes), enquanto que outros resultaram da leitura da cena pelos participantes através de suas específicas perspectivas, como é o caso da forma de jogo. Mesmo um traço como aparência, que à primeira vista (inevitável jogo de palavras) é tão evidente, foi interpretado e julgado de diferentes maneiras; por vezes foi referido a qualidades ontológicas ("virtual" em oposição a "real"), por vezes como uma percepção de traços disponíveis para avaliação da atitude de personalidade ("*expressões faciais, alegria, tristeza, pensativo*"), por vezes simples apreciação estética ("*bonito*", "*mais feio*").

Os critérios e os atributos citados pelos respondentes, e a maneira como são aplicados, formam um conjunto coerente. No grupo em que um atributo era mais explicitamente destacado – aparência – este atributo, ou atributos que a ele remetam, foram sistematicamente colocados como "semelhanças". Complementarmente, outros atributos – em especial personalidade – foram citados como diferenças, como resultado da busca, dentro da referência dada pelos dois vídeos, daquilo que pudesse ser interpretado como diferença, dado que visualmente ambos os agentes eram similares. Neste contexto, os resultados variados obtidos para a questão "é o mesmo ou não é" refletem a consideração produzida por cada sujeito, em resposta a uma pergunta que mostrou-se extremamente provocativa. Afinal, se os agentes *parecem* visualmente os mesmos, então *deveriam* ser os mesmos; se há um questionamento sobre se são os mesmos leva os participantes a procurar a diferença não-óbvia, aplicando as competências que estes participantes têm no uso cotidiano de computadores e desta forma de mídia comunicativa para descobrir esta distinção oculta.

O grupo do questionário "Mesma personalidade" pode ser compreendido em comparação com o grupo anterior. Quando os agentes compartilhavam não a prontamente visível aparência, mas uma atitude, a personalidade, o que se verifica é que a pergunta

também se torna provocativa. Se os agentes obviamente não são visualmente a representação um do outro, e há o questionamento sobre a identidade de ambos... então também a semelhança deva ser procurada, colocando em ação as competências já mencionadas. Atributos de aparência são trazidos para argumentar como semelhança e também como diferença, atributos de atitude, personalidade, também são esmiuçados e interpretados ora como semelhança, ora como diferença. Em um caso como outro, as instâncias do agente são compreendidos a partir de traços observados que são interpretados como atributos, e estes atributos dão suporte à compreensão, na forma como ocorre e é relatada, do agente.

Como pode ser observado das respostas dadas, não era difícil para os participantes perceber a sequência de eventos apresentado em cada vídeo como trocas relacionais que ocorriam entre uma pessoa e uma individualidade consistente e persistente. A consistência e persistência podem ser vistas na designação de atributos de "personalidade" e de processos temporais à sequência de comportamentos dos agentes artificiais, daí construídos como "personagens". Embora a leitura e interpretação da sequência de condutas das figurações visuais pareça um tanto óbvia, é de fato notável no sentido de que robôs e desenhos animados não são pessoas e não há expectativas de que sejam ativamente relacionais, no sentido social.

O trabalho interpretativo realizado pelos participantes cria, a partir das cenas observadas em vídeo, um mundo estável e estruturado, um mundo significativo em termos sociais e psicológicos. Traços observados foram tomados como pistas, ou indicações, que descrevem e comprovam uma estrutura subjacente, e esta estrutura subjacente é por sua vez recursivamente trazida para dar significado a evidências individuais (Garfinkel, 1984, chap. 3). Por outro lado, a facilidade no reconhecimento de personagens não tornou mais simples a decisão, a partir dos recursos disponíveis, se os dois personagens apresentados eram o mesmo ou não. Os participantes "consultaram características institucionalizadas da coletividade como um esquema de interpretação" (1984, p. 92). Os esquemas disponíveis para os respondentes eram relacionados a ontologias compartilhadas na comunidade envolvida, isto é, que tipos de entidades existem e como é possível relacionar-se com elas. A "agência" dos "agentes" não foi colocada em questão, mas o trabalho de *identificação* foi dificultado porque processos socialmente sancionados para identificação (isto é, como reconhecer uma entidade

reencontrada como sendo a mesma, em diferentes situações) não estavam presentes. Pode-se mesmo argumentar que estes processo, para este tipo de entidades, nem sequer existem ainda. De qualquer forma, traços intencionalmente construídos não são suficientes para prover a identificação; tais traços são propostos e tidos como significativos pelos projetistas, mas não há garantias de que conduzirão (se conduzirem) aos mesmos processos de interpretação quando o artefato for explorado por usuários. De maneira similar ao que ocorre em relações entre humanos, há a possibilidade de não serem notados, de interpretações "equivocadas" (isto é, em desacordo com o significado originalmente pensado para o traço, pelos projetistas), e de pessoas criaram significados inteiramente novos. É interessante observar que, como visto pelas respostas que destacaram o "estilo de jogo" dos personagens, é esta mesma indeterminação e abertura que podem vir a ancorar a interpretação de artefatos como entidades dotadas de agência.

Ao ser abordada desta forma, percebe-se como a interação comunicativa e social entre humanos e não-humano vai ocorrer em um contexto sempre aberto a ser recursivamente empregado e apropriado como recurso, e é ao explorar esta característica que o humano aporta riqueza e interesse às suas relações (Suchman, 2007, p. 67 e 81). Ao focar a atenção sobre o utilizador, vemos em curso o processo pelo qual o humano torna o personagem artificial não apenas inteligível, mas capaz de propor sua inteligibilidade – em outras palavras, o processo de propor o personagem como agente inteligente.

## *6.6 Considerações sobre o caso*

A situação que aqui destacamos tem como centro um artefato, um conjunto de utilizadores, e dois pesquisadores: foi investigada a forma como pessoas recrutadas em um ambiente técnico com convivência cotidiana com computadores encontram um sistema específico de inteligência artificial construído dentro do grupo de IA, e como eu e Gepê, enquanto pesquisadores, fomos levados a nos interrogar sobre como compreender um resultado inesperado. A noção explicativa central que surgiu foi a atividade interpretativa realizada pelos usuários. Observamos como os usuários, participantes oriundos de uma universidade técnica e de fóruns técnicos na internet, exerceram um repertório de

estratégias para interpretação e raciocinaram sobre o material audiovisual a eles submetido. O resultado dos julgamentos a eles solicitados não foram uniformes, distribuindo-se amplamente entre o reconhecimento e o não reconhecimento da "identidade" de um agente quando presente em um dispositivo ou outro. No entanto, os critérios utilizados e os atributos considerados importantes foram consistentemente levantados e comentados pelos utilizadores.

Estas observações remetem às teorias de ação situada e do método de interpretação documentária. Estas teorias salientam o caráter não determinado do mundo observado, onde a inteligência humana vai discriminar elementos, como indícios interpretativos, a partir de teorias e categorias explicativas de mundo, e vai recursivamente aplicar a existência e articulação destes indícios como evidências de suporte para as teorias e categorias. Em outras palavras, os utilizadores aplicaram seu conhecimento e sua capacidade de discernimento a uma situação para eles ainda não organizada cognitivamente, e o que observamos foram os critérios com os quais procuraram dar ordem e compreender, tornando inteligível, a situação a eles apresentada. Os atributos percebidos pelos usuários não se esgotaram naqueles selecionados pelos pesquisadores para gerar a performance dos agentes, mostrando como ao ser interrogado pelos sujeitos estas performances podem ser vistas e interpretadas de maneiras novas e surpreendentes.

Em suma, o que propomos aqui é uma perspectiva para o projeto mais amplo da Inteligência Artificial, que visa enfatizar a rica subjetividade produzida na relação dos sujeitos que interagem com sistemas com caraterísticas humanas. Propomos a observação atenta de como estes sujeitos agem e da forma como descrevem suas ações, dando conta de uma subjetivação que se mostra mais complexa do que um conjunto aparentemente claro de "benefícios educacionais" discretos e catalogáveis. Sugerimos também, em função dos resultados obtidos, que esta pode ser uma perspectiva frutífera e informativa de análise deste tipo de interação, abrindo a possibilidade de novos insights sobre as relações afetivas e sociais entre humanos e máquinas. Em particular, enfatizamos que decursos inesperados na interação entre humanos e artefatos *inteligentes* possam ser considerados como geradoras de novas, não previamente projetadas, perspectivas, ao invés de tratadas como problemas ou falhas. Nosso argumento é que a abrir a possibilidade de novas formas de ver as relações sociais e afetivas em que humanos e artefatos tecnológicos participam é uma forma de tornar a tecnologia mais humana.

# 7 Agências entre sistemas computacionais e humano

Este capítulo pretende explorar temas que se relacionam com o conceito de agência a partir de algumas de suas implicações na proposição, pelos participantes dos grupos onde foi realizado o trabalho de campo, de sistemas computacionais a que dão o nome de agentes artificiais ou agentes inteligentes. Para a IA, agentes têm uma definição muito específica, cuja intenção é guiar a produção de sistemas através de uma certa estratégia de projeto e desenvolvimento de programas (software). No entanto, esta definição é atualizada em sistemas que ancoram noções específicas sobre como pessoas articulam-se e agem ao utilizá-los, e ramifica-se em diversas noções do agir humano e das capacidades de objetos tecnológicos que são propostas pelos pesquisadores de IA. Mais especificamente, abordaremos aqui a pesquisa, pelo grupo, com o objetivo de produzir agentes artificiais aptos a negociar *emoções* e *sociabilidades expressivas*. Vista por este viés, a proposta de agentes artificiais reverbera de diversas formas no encontro com outras concepções de agencialidade.

Uma das maneiras de expressar esta definição, retomada de uma obra de referência da IA, utilizada pelos participantes do grupo, poderia ser "algo que possa ser visto como percebendo seu ambiente [...] e agindo sobre esse ambiente" (Russell & Norvig, 1995, p. 32). Do ponto de vista da técnica computacional, podemos nos questionar em que um sistema projetado a partir da estratégia de agentes inteligentes é particularmente distinto de outras abordagens de projeto para sistemas computacionais. A resposta poderia ser a recomendação de que a estratégia baseada em agentes (Jennings et al., 1998; Luger, 2002, p. 236) produza sistemas compostos por diversos componentes (programas), cada qual direcionado a um aspecto do problema, de maneira autônoma, no sentido de prover por

suas ações de interação com outros componentes, e flexível, significando que deve apresentar ações que incluam respostas alternativas em momentos apropriados. Segundo estes autores, requer-se ainda que sejam situados, ou seja recebendo entradas e agindo diretamente sobre o que constitui seu entorno ambiental, e ainda sociais, interagindo apropriadamente com, outros softwares e com pessoas. É interessante observar que requisito "situado" aproxima esta concepção daquela de robótica situada (Brooks, 1995), aproximação esta que não é incidental, sendo explicitada por (Luger, 2002, p. 801).

A conceituação de agente pode ser ainda mais sintética – e mais abrangente – como a colocada por Russel (2010, p. 34): "qualquer coisa que possa ser vista como percebendo seu ambiente através de sensores e atuando sobre este através de atuadores". A definição, abrangente e inclusiva, é expressamente aproximada do humano: "o agente humano tem olhos, ouvidos... como sensores e mãos, pernas, cordas vocais... como atuadores", embora Russel vá enfatizar, ao longo do livro, o desenvolvimento de estratégias de projeto de agentes artificiais baseados em computadores. Esta abordagem é presente no grupo brasileiro, o que pode ser percebido em falas da professora líder e em materiais de instrução produzidos por ela. As abordagens de Russel e Luger são similares entre si e a outras presentes na IA, enquanto estratégia de desenvolvimento de software; a diferença que nos interessa é a abordagem de Russel aproximar, de maneira explícita apresentada, as agências humana e artificial. Em comum, ambas as conceituações enfatizam a delimitação do agente, unidade individual e isolada-conectada ao mundo "exterior" através de canais específicos de entrada e de saída.

A abordagem de agentes própria da IA é cognitivista no sentido de considerar operações mentais como mediadoras entre os estímulos do ambiente e a resposta do indivíduo em forma de comportamento (Suchman, 2007, p. 37). Estas operações mentais seriam então passíveis de abstração – e de serem realizadas em máquinas computacionais. A cognição, nessa perspectiva, não é apenas comparável à computação; para a abordagem cognitivista, cognição *é* computação. A inteligência e os fenômenos mentais são localizados no cérebro, internos ao corpo e individuais. É justamente esta forma de abordar a questão que torna a inteligência candidata viável à sua replicação em máquina. O "agente genérico" é a formulação, segundo a inteligência artificial, da entidade autônoma e individual que age sobre o ambiente que se localiza do lado de fora, através de canais específicos de entrada e saída de informação. Este agente genérico pode ser

tanto um humano, modelo da agencialidade, como um ente artefatual, computacional, capaz de conduta semelhante. Ver Figura 1 (Russell & Norvig, 1995, p. 32).

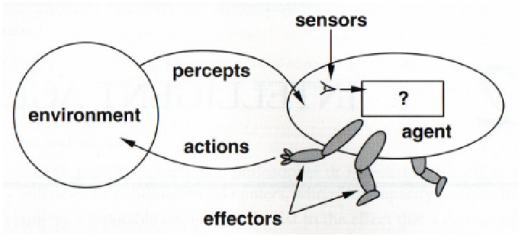


*Figura 1*

A noção de agente apresentada pela IA pode ser interrogada a partir das afirmações de afinidade com o humano e com o social desta agência. A formulação deste agente, por Russell e Luger, e também visível na definição que os participantes dão ao desenvolver agentes como seu trabalho, e separa claramente o agente do meio, e o caracteriza como o que age sobre o meio. Agência, segue-se, é uma propriedade abstrata, que se atualiza em instâncias de agentes existentes. A putatividade da agência de seus artefatos é prioritária para a IA, que menciona explicitamente o referencial humano como inspirador de seu projeto. Há então uma questão implícita a ser abordada quando se está tratando de uma agência potencialmente atribuível como a da máquina; isto é, em diferença à agência tacitamente considerada do humano.

A atribuição de agência, no entanto, não é ponto pacífico na avaliação que se faz do mundo e o que nele existe, isto é, nas ontologias consideradas por pessoas. A agência que se propõe para o artefato parece claramente definida a curta distancia no momento em que se está descrevendo o conceito orientador para a construção de um tal artefato ou enquanto um artefato real está sendo construído. A uma distância maior a problemática passa a ser a de dar conta da agência potencial atribuída ao ente. A questão que é colocada é a seguinte: se retrospectivamente uma capacidade para agir é reconhecível e designável ao candidato a agente. Neste ponto, seguimos o questionamento de (Fuller, 1994), que aponta como traço discriminador para dar conta de uma determinada agência a interrogação sobre diferença: sabemos que a ação do ente é relevante se a interveniência desta ação foi determinante para o curso de ocorrência observada, e portanto constituiu diferença neste curso. A conclusão seria a de que, neste caso, a atribuição de agência faz sentido. Fuller então aponta que, se utilizarmos o humano como referência, é preciso

reconhecer que a agência não é uma característica homogeneamente distribuída mesmo entre humanos: ao relatar retrospectivamente a História, historiadores conferem a determinados humanos-agentes a característica de que foram responsáveis por eventos – implicando uma agência menor de outros tantos, que, segundo esta narrativa, poderiam ter agido diferente sem fazer diferença.

Bruno Latour, ao propor a Teoria Ator Rede (ANT, Actor-Network Theory), traz para o centro da discussão exatamente a questão da atribuição de agência (Latour, 1994). O ponto de partida é semelhante à consideração sobre agentes proposta por Russell: uma simetria ao tratar tanto de humanos como de não-humanos, frente ao problema da agência. A diferença está em que Russell, e os cientistas da IA que trabalham com agentes artificiais, partem de uma premissa ontológica, a de que o humano *é* agente, sendo essencialmente o agente modelar, e de que entes artificiais serão agentes na medida em que adquirirem certas capacidades demarcadas. Latour, por outro lado, inspirado no princípio de simetria proposto originalmente no âmbito da Sociologia do Conhecimento (Bloor, 1991), propõe tratar humanos e não-humanos de maneira simétrica com respeito ao problema da agência, procurando não a atribuir a priori nem a um conjunto de entes, nem a outro. A pergunta, "Quem age?", é dirigida, pela ANT, de maneira imparcial a humanos e não-humanos.

As respostas encontradas pela ANT a esta pergunta têm sido várias e interessantes (Callon, 1986; Latour, 2001). Ao invés de uma agência humana *ex nihilo*, isto é, primária, definitiva e modelo a ser seguido por outras agências postulantes, os autores ligados a esta teoria encontraram o exercício do "fazer diferença no curso da ocorrência" distribuído e arranjado de múltiplas maneiras. A estes arranjos, em que as capacidades para agir acontecem na negociação entre conjuntos de humanos e não-humanos, Latour chama de atores-rede (ator, neste caso, remetendo à característica de agir, não à de representar um papel). Observaram, por exemplo, que ao estabelecerem formas de agir com o concurso de novas tecnologias, as pessoas em geral não se concentram em categorizar previamente se a agência que está sendo criada pertence à máquina ou à pessoa. A nova agência é simplesmente colocada em prática. Os entes do mundo ganham uma caracterização, uma identidade, no encontro com as pessoas que se interessam, ou que se colocam no caminho destes entes. Estes entes oferecem uma resistência, própria de seu ser ao encontrar-se com as pessoas – esta resistência é agência, é a forma própria destes entes agirem. Em (Callon,

1986) é apresentado o caso dos cientistas cujo projeto era repovoar a Baía de St. Brieuc (localizada no noroeste da França) com a espécie dali nativa de mariscos, naquele momento em risco de extinção por sobrepesca. Para o projeto ser realizado, era necessário contar com a colaboração dos pescadores – para que não pescassem os mariscos na área experimental. Era também necessário compreender o ciclo de vida dos mariscos, para que o processo de incentivo à reprodução dos mariscos, que estava sendo desenvolvido, chegasse a termo. O que Callon destaca é que, enquanto o circuito cientistas-mariscos-pescadores não estava estabelecido, o processo não era simplesmente estabelecer regras prévias para o comportamento que os pescadores deveriam seguir. O processo envolvia negociar com os pescadores, dentro do que era conhecido e do que era desconhecido também sobre o marisco, persuadindo-os a respeitar o acordo mesmo sem conhecimento suficiente sobre o processo para ter garantia do desenlace. O processo envolvia, adicionalmente, agir de maneira tentativa e procurando interpretar as evidências esparsas que eram coletadas sobre os mariscos, na expectativa de que este curso de ação obtivesse resultados que pudessem ser considerados adequados pelos pesquisadores e pelos pescadores. Em outras palavras, os cientistas estavam negociando também com os mariscos, reconhecendo e adequando-se a estes agentes peculiares, sem que isto significasse antropomorfizar esta agência.

A teoria ator-rede auxilia a compreender também o fenômeno da agência delegada, e daquela embutida na infraestrutura técnica e social, agências estas apagadas por graus de invisibilidade. A agência delegada refere-se ao ente que é desenhado para desempenhar uma função originalmente designada a uma pessoa; Latour exemplifica (Latour, 2001) com o caso da lombada – adequadamente chamada de guarda-deitado em alguns lugares, que materializa e transfere a agência do guarda de trânsito no que se refere à diminuição de velocidade no trecho em que foi construída. O motorista que diminui sua velocidade em função da lombada não é remetido à autoridade do guarda, nem precisa refletir sobre a lei de trânsito e a velocidade permitida naquele trecho; mas o objeto físico produziu a ação, prevista no plano da autoridade de trânsito. Agências também são embutidas na infraestrutura com que lidamos diariamente. Aqui, Latour exemplifica (Latour, 2001, p. 211) com uma aula em que o professor utiliza um retroprojetor. Durante a aula, o agente visível, produzindo ação com propósito, é o professor em conversa com seus alunos; mas no instante em que o dispositivo falha e interrompe o curso desta ação, seu papel é

ressaltado, e, em função da diferença ocorrida no curso da ação, a participação do dispositivo é visibilizada. Quando ocorre a falha, a infraestrutura torna-se aparente, e a ação pode ser então apreendida como atribuível ao conjunto professor-retroprojetor, reconhecendo a participação deste último sem, no entanto, tomá-la como equivalente à do humano. A agência invisível torna-se aparente neste instante.

O argumento colocado acima é, à primeira vista, pouco útil na medida em que não estabelece um conceito claro, em forma de critério, para distinguir tipos de agência: o centro do argumento está em tornar visíveis muitas formas de agência, sem no entanto, inicialmente, insistir em separá-las. A teoria ator-rede, no entanto, é precisamente uma maneira de abordar a questão da agência a partir da observação de seus efeitos e dos entes que estão envolvidos, e estimulando a descrição dos agentes dentro das configurações dentro das quais agem. O que pode ocorrer é que, examinando-se as peculiaridades de cada arranjo técnico e social que configura agências específicas, o quadro do humano como modelo de agência passe a complicar-se, deixando de haver uma correspondência clara entre diversos níveis de agência abstraída (Lee & Brown, 1994) – mas a sensibilidade ao arranjo local (e situado, no sentido da ação situada de Suchman) é precisamente uma das características da ANT.

O que emerge da discussão a respeito da agência, realizado pela ANT e também por outras abordagens (Ashmore et al., 1994), é uma perspectiva em que a agência é problematizada, estando distribuída em arranjos (redes, segundo a ANT) em que co-participam humanos e não-humanos (isto é, os objetos técnicos trazidos para auxiliar a compor estas redes). A agência dos objetos é considerada de maneira simétrica como princípio de estudo, mas isto não significa que seja equivalente à humana. E as diversas agências não são sempre visíveis ou reconhecidas; pelo contrário, quando o arranjo apresenta um problema, o papel desempenhado por seus componentes é trazido para a visibilidade, e é possível então perceber como as ações não estão concentradas em um "agente total", mas habilmente distribuídas dentro do arranjo.

Tendo colocado em perspectiva duas abordagens diferentes para a compreensão do conceito de agente, exploraremos a seguir alguns momentos em que os participantes do grupo desenvolvem seu trabalho e produzem objetos técnicos sob esta denominação. Buscaremos compreender a produção destes sujeitos, realizada dentro da orientação de

sua concepção de agente, e como compreender estes objetos enquanto agentes dentro desta proposta. Também procuraremos relacionar esta produção com a noção de agencialidade instituída dentro das redes sociotécnicas observadas, examinando como se posicionam, dentro destas redes, os artefatos construídos pelos participantes: os agentes artificiais.

## *7.1 Afetos e compreensões*

> Suzane está à mesa, e tem diante de si um tabuleiro grande, com 0,5m de largura, contendo um conjunto de peças identificadas para o jogo da velha – algumas marcadas com um "○", outras com um " ×". Do lado de lá do tabuleiro, há um boneco robô, "fofinho", um Gato amarelo com os olhos bem abertos. Ao lado dela está Gepê, meu colega de pesquisa. Suzane olha para mim, depois para meu colega, com um ar interrogativo. Eu olho para Gepê, que inicia então a apresentação da atividade a ser realizada: resumindo, Suzane deveria jogar o jogo da velha com o Gato, e depois iríamos entrevistá-la sobre sua experiência.
>
> Gepê concluiu a apresentação da atividade, e o Gato então levanta a cabeça e fala. A sua voz é acompanhada de movimento das sobrancelhas e dos lábios! Convida a interlocutora à sua frente: "Vamos jogar o jogo da velha?". Suzane faz um primeiro lance, e o Gato então fala o seu lance. As jogadas se sucedem, o Gato ocasionalmente fazendo algum comentário. Suzane vai jogar algumas partidas com o Gato.

Assim como Suzane, uma série de outras pessoas participaram deste e de outros estudos de interação com este Gato animado. Quase todas as sessões com o Gato seguem um roteiro do tipo "jogar um jogo de tabuleiro, em lances alternados, com falas do Gato relacionadas ao jogo". O interesse central aqui é ver as pessoas frente a um boneco, um robô "animado", dotado de meios expressivos tais como voz e também lábios, sobrancelhas e olhos que se movem, e observar a experiência destas pessoas diante da presença e da performance do robô.

O programa de computador que controla os movimentos e a animação do robô, sua fala e seu jogar é um agente, um "agente artificial". No grupo, cada projeto de estudo com

o Gato programa seu próprio agente, embora partes do programa sejam rotineiramente reutilizadas de um projeto para outro. O Gato é um equipamento muito útil para os estudos com agentes que o grupo realiza; parece-se com um personagem de desenho animado, fala e move-se, e as pessoas divertem-se ao experienciar uma sessão de atividade com ele. Os agentes são programados, por sua vez, de acordo com as características que deseja-se estudar. Assim, é possível para os participantes do grupo a observação do encontro de pessoas com o Gato, modulada pelo agente que estiver sendo executado no momento.

A ideia promovida pelo grupo é criar agentes com características emocionais e sociais, para auxiliar ou entreter pessoas em tarefas. O objetivo é levar as pessoas a comportarem-se de maneira mais natural ao utilizarem computadores, tornar mais fácil e mais atraente o uso do computador. É interessante observar como, nos diversos estudos realizados pelo grupo, é comum que ao avaliar sua experiência com o Gato pessoas utilizem referências explícitas a termos relacionados a personificação, emoção e sociabilidade. Em uma das sessões experimentais de que participei, usuários que jogaram contra o Gato referiram-se ao seu "humor, o sarcasmo", disseram que parece "simpático [...] delicado em suas intervenções", e admitiram perceber "algumas expressões [nos traços fisionômicos móveis do robô] no Gato, não sei se foi de propósito ou não".

O trabalho do grupo, em torno de agentes com características emocionais e sociais, insere-se dentro de um campo de pesquisa muito ativo e produtivo da computação. Enquanto que a ciência de programar computadores não parece, à primeira vista, território ameno para dar conta de emoções, os participantes do grupo não parecem intimidar-se com o desafio, no entanto, e fazem mesmo disto o seu cotidiano. A emoção, e sua abordagem dentro da informática, a computação afetiva, são conceitos que ocupam um lugar central dentro da produção do grupo, sendo a emoção considerada uma chave essencial para lograr o objetivo que impulsiona o seu trabalho, que é a de produzir agentes artificiais como humanos, marcantes e que despertem interesse[11].

O trabalho do grupo liga-se, neste âmbito, ao da computação afetiva. A maneira de formular os problemas, e as formas de resolvê-los, no que tange à relação considerada como afetiva de humanos com máquinas, segue e refere-se com frequência aos projetos da

11 Ver a discussão sobre a delimitação de termos relacionados ao afetivo na p. 40.

computação afetiva, integrando-os ao marco de referência de agentes artificiais. Agentes artificiais, na pesquisa do grupo, são tornados com emoção em um processo que inicia com premissas de trabalho, expressas nos meios de apresentação do grupo e de seus referenciais teóricos – nas páginas internet, em trabalhos introdutórios, em materiais de ensino, e em momentos em que membros são apresentados aos temas de investigação do grupo. Estas premissas iniciais referem-se, então, a emoções como uma característica muito importante para a conduta humana, e levar esta característica em conta é uma maneira de aprimorar agentes artificiais como artefatos tecnológicos.

A partir destas premissas são propostas maneiras de lidar, computacionalmente, com a emoção, ao mesmo tempo estabelecendo o que conta como emoção para os participantes no contexto da relação de seus produtos com pessoas. Programar a demonstração ou expressão de um estado emocional subjacente é, singularmente, uma das formas de dar ao agente esta característica emocional. Por outro lado, reconhecer ou identificar um estado emocional no utilizador também é considerado importante, e, por fim, encaixada entre estas duas capacidades, soma-se a de modificar sua conduta em função do estado emocional próprio ou do utilizador humano. Modelar os entes participantes deste circuito a partir do modelo do agente da inteligência artificial é produtivo do ponto de vista da produção de sistemas computacionais inteligentes, e é possível notar um cuidadoso acoplamento entre as noções de humano, de agência, de afeto e de comunicação que são empregadas. No modelo da IA, de agente como ente atômico e insular que acessa o seu exterior através de canais definidos, a noção de afeto como um estado interno definível acopla-se adequadamente. A possibilidade de reconhecer a emoção refere-se, neste modelo, ao identificar do fenômeno que ocorre dentro de uma faixa de variação previamente conhecida e que manifesta-se localizadamente nos canais de saída do agente. E, completando o circuito, o reconhecimento da emoção baseia-se no apresentar desta manifestação aos canais adequados de entrada do agente.

A descrição do processo é paralela, ou simétrica, entre o humano e o sistema computacional: é viável o desafio de projetar o agente artificial que reconhece emoções porque o agente humano, nesta perspectiva, dispõe de um estado interno, que se manifesta através de um código legível e coerente em canais específicos e disponíveis à leitura. O humano por sua vez reconhece claramente o estado emotivo do agente artificial,

porque este corresponde a um item conhecido e compartilhado do catálogo emotivo que foi expresso – e feito legível – nos canais apropriados do agente.

### 7.1.1 Computação afetiva e emoções no balanço da legitimidade

A IA dos agentes artificiais vai buscar à computação afetiva o programa de trabalho para pensar e projetar, em forma computacional, esta proposta de relação humano-máquina. Na computação afetiva, como praticada e sustentada pelo grupo – e também pelos participantes do grupo no Brasil – é considerado natural desenvolver esta proposta a partir de uma teoria psicológica sobre o emocional, isto é, a partir de um conhecimento especializado, e autorizado, sobre o tema. Uma das abordagem preferidas pela computação afetiva é a teoria de emoções exposta em Ortony et al. (Ortony et al., 1990): uma teoria *cognitiva* das emoções. A teoria de Ortony et al. é conhecida como Teoria OCC (a partir do nome dos autores: Ortony, Clore, e Collins), e é citada amplamente pelos trabalhos de IA que mencionam afeto computacional. O interesse é centrado em torno de um esquema estrutural para emoções, proposto pelos autores (ver Quadro 3). Segundo os autores (pp. 2-4), emoções são reações de avaliação ou juízo, positivo ou negativo, em relação a situações percebidas pela pessoa. Há uma cadeia causal definida, e emoções surgem como resultado de um processo cognitivo; efeitos fisiológicos, comportamentais e expressivos são posteriores, pressupondo um primeiro passo cognitivo. Seguindo estas premissas, é possível focar em distinto tipos de emoção, vinculados a certas condições de produção da emoção, ao invés de abordar uma multiplicidade indeterminada de estados emocionais discrimináveis.

As condições de produção são, portanto, apresentadas por Ortony et al. como a avaliação das consequências de eventos – para si, ou para outros – ou como avaliação de ações praticadas (por agentes) ou, ainda, por avaliação de aspectos de objetos. A determinação do campo reativo à combinação destas condições o divide em um número de emoções definido para a teoria OCC, que são 22. Este número, os autores explicam, correspondem à cardinalidade de uma *tipologia* das emoções, a partir das condições; o número de palavras existentes para descrever "estados emocionais" é maior, mas elas corresponderiam a especificações *dentro* dos tipos, não a situações diferentes. A teoria, é ressaltado, é sobre *emoções* em si, não sobre as palavras que as descrevem. O foco

ontológico, substantivando a emoção, é evidente em outras distinções que os autores procuram deixar claras, como a de que a "fisiologia é essencial para a experiência emocional mas não é relevante para a questão do papel exercido pela cognição na produção de emoções", ou ao afirmarem que o objetivo é uma "descrição lógica, não temporal" da emoção. A teoria é desdobrada pelos autores em torno da lista das reações a serem presumidas, explicadas no contexto das situações de produção. Um quadro de referência explicativo é traçado para uma psicologia da avaliação, para a questão da intensidade das reações, e para as condições que dão origem a reações avaliativas, positivas ou negativas – em outras palavras, as emoções da teoria OCC: eventos, ações, percepção de objetos. O conjunto de emoções considerado vai sendo composto à medida que são exploradas estas situações, e os julgamentos presumíveis para cada uma destas são trazidas e examinadas.

A ampla adoção desta teoria pelas comunidades da Computação Afetiva e da IA é em si um fenômeno de interesse. Em primeiro lugar, a teoria é prezada pela sua simplicidade organizacional, e pela sua objetividade causal: é uma teoria *implementável* em termos computacionais, segundo os participantes. Este é um argumento utilizado em momentos em que participantes com mais experiência apresentam esta teoria a novos participantes, como palestras ou aulas, mas, mais notavelmente, é um argumento que revela-se consistente quando os novos participantes enfrentam a tarefa de desenvolver seus próprios sistemas em que lidar com variáveis de emoção é um requisito. A explicação "OCC é uma teoria implementável" foi apresentada a mim também por participantes alunos, que desenvolviam seus próprios projetos, e que a utilizavam para construir as funcionalidades destes projetos.

Esta adequação para a utilização na computação, no entanto, não é gratuita. Pelo contrário, é parte constituinte do desenvolvimento da teoria, um requisito de projeto: "gostaríamos de estabelecer os fundamentos para uma teoria computacionalmente tratável da emoção" (p. 2). Os autores explicitam a intenção de criar um quadro teórico traduzível e derivável como formalismo computacional (p. 181), e antecipam a possibilidade de estabelecer um conjunto de regras lógicas apropriáveis por sistemas de IA com o objetivo de raciocinar sobre estados emocionais. Estas regras teriam como antecedentes representações abstratas de situações, como as examinadas ao longo do livro:

"expectativa", "desejo", e como consequências o "potencial para emoção", ou ainda "intensidade da emoção".

A cuidadosa construção da teoria dá ao trabalho de Ortony et al. uma qualidade de objetividade e coesão, além de uma organização interna entre os construtos teóricos que a torna atraente como ponto de partida para uma modelagem funcional, computacional, baseada em processamento de informação. No entanto, estas características têm algumas contrapartidas, que tornam-se visíveis ao examinarmos o procedimento pelo qual, pouco a pouco, os autores montam seu quadro de referência sobre emoções.

Em primeiro lugar, destaca-se a forma como o fenômeno emotivo é construído como objeto em si, destacado tanto do corpo do sujeito onde se produz a emoção, como também do coletivo em que este sujeito realiza sua vida e suas experiências emocionais. Ambos, o corpo e o coletivo, são empurrados para um segundo plano, em que figuram estaticamente como "contextos" dados. A cena é ocupada, primariamente, pela emoção ontológica, que assegura seu *ser* nas regras de que é a consequência. Neste sentido, a teoria de Ortony et al. alinha-se a uma tradição de pesquisa em emoção que paulatinamente produziram um fenômeno analisável, quantificável e discriminável, através de procedimentos investigativos que enunciavam-na como um fenômeno inteligível independentemente do corpo. Esta tradição é analisada por Dror (Dror, 2001), que mostra como foi construída ao longo do tempo, a partir do final do século XIX, produzindo, entre outros aparatos que falam tecnicamente sobre o corpo que sente, o detector de mentiras, e os "testes de paixão", encartados em revistas populares, em que um cartão sobre o qual o leitor pressionava o dedo mudava de cor de acordo com seu "estado emocional" (apaixonado ou indiferente).

Esta ontologização e colocação em cena de uma emoção-objeto, independente e dotada de uma existência própria, é efetuada, dentre outras estratégias de produção destes objeto do saber, pela definição através de regras que a determinam. Esta definição ocorre de uma maneira sistemática, ao longo da obra, através de um método específico de argumentação que iremos rever a seguir.

As situações que os autores consideram que dão origem a reações do tipo emoção são apresentadas, descritas e especificadas, resultando em um campo emotivo completamente preenchido e analisado. Ao apresentar, por exemplo, as emoções relacionadas ao

julgamento sobre eventos, são definidas as possíveis situações: um "EVENTO DESEJÁVEL" e seu reverso, um "EVENTO INDESEJÁVEL". A desejabilidade, neste caso, é elaborada como dependente de fatores próprios, ou *contexto*; desejabilidade, portanto, deve ser "computada em um contexto" (p. 89). A intensidade da emoção resultante, por sua vez, também tem seus fatores, que são a própria desejabilidade, e *variáveis globais*, como inexpectativa e proximidade (temporal do evento). De maneira similar são apresentadas as situações relacionadas a julgamentos sobre ações (de agentes) e sobre aspectos de objetos. As emoções propostas, então, são relacionadas a cada uma das situações definidas. Assim, à situação "EVENTO DESEJÁVEL", a emoção-tipo associada é a de "contentamento", à situação "EVENTO INDESEJÁVEL", a de "descontentamento". Quando a situação é avaliada em relação a outrem, temos então "EVENTO DESEJÁVEL" dando origem a reações de "contentamento" e "descontentamento" (tais como ressentimento), e "EVENTO INDESEJÁVEL" dando origem a reações de "descontentamento" (como tristeza solidária) e contentamento. A construção é iterada para cada uma das outras emoções-tipo propostas; como o julgamento sobre "AÇÕES DE SI, LOUVÁVEIS", "AÇÕES ALHEIAS, LOUVÁVEIS", ou sobre objetos, como as reações a "OBJETO ATRAENTE" ou "OBJETO NÃO-ATRAENTE" e assim por diante.

<table>
<tr><td rowspan="6">Consequências de eventos:</td><td rowspan="2">Consequências para outrem:</td><td colspan="2">Desejável para outrem: feliz-por-outrem, ressentimento</td></tr>
<tr><td colspan="2">Indesejável para outrem: feliz-por-desgraça-alheia, piedade</td></tr>
<tr><td rowspan="3">Consequências para si:</td><td rowspan="2">Relacionado a expectativa:</td><td>Confirmação: satisfação, receio-confirmado</td></tr>
<tr><td>Desconfirmação: alívio, desapontamento</td></tr>
<tr><td>Não relacionado a expectativa: contentamento, descontentamento</td><td rowspan="3">Compostos: gratificação, remorso, gratidão, ira</td></tr>
<tr><td colspan="2">Ações de agentes:</td><td>Ações de si: orgulho, vergonha</td></tr>
<tr><td colspan="2"></td><td>Ações de outros: admiração, reprovação</td></tr>
<tr><td colspan="4">Aspectos de objetos: amor, ódio</td></tr>
</table>

*Quadro 3: Emoções e situações antecedentes na teoria OCC*

Em resumo, as emoções são definidas como a reação, ou "positiva" ou "negativa", presumidas ocorrerem em resposta a estas situações. O campo situacional é pensado e proposto de maneira a esgotar as possibilidades emotivas (embora, sempre ressaltem os

autores, há muitas palavras relacionadas a emoção, que significam na verdade uma posterior especificação de alguma das 22 emoções-tipo).

Propor uma análise tão ordenada, abrangente e especificada da vivência emocional, no entanto, vai levar a algumas dificuldades em tratar da diversidade da experiência humana. Um caso claro ocorre quando autores interrogam: "Podem pessoas genuinamente experimentar gratidão em relação a alguém que apenas acidentalmente efetua um desfecho/resultado desejável para elas?"(p. 150). A resposta ocorre deduzindo que as *condições* necessárias para a gratidão, dentro do quadro teórico, não são satisfeitas – isto é, não houve intencionalidade do agente implicado no desfecho. E a resposta é dada com certeza: "Nossa resposta é não, e a única razão concebível para crer diferentemente é que pessoas às vezes *falam* desta maneira" (i.e. expressando uma gratidão que não sentiriam). É notável nesta perspectiva o fechamento dado, previamente, à experiência humana; categorizado assim *a priori*, não há espaço para um sentir do sujeito, porque o fenômeno é definido em termos de regras. No sistema coerente de regras, sentir de outra maneira não é sequer transgressão, é *inexistente* porque não acorda com a definição. Caso o sujeito insistir, isto é desqualificado como um relato falsificado, cuja culpa é da linguagem e da convenção social.

A teoria OCC procura, desta forma, assegurar a todo custo uma *estabilidade* da emoção por via da definição, "uma descrição lógica, não temporal" (p. 19), isto é, uma emoção purificada de seu caráter processual e histórico marcado na história do corpo que sente. Este é o motivo pelo qual ao descreverem a emoção de jogadores e torcedores em eventos esportivos (p. 4), sua conclusão é de que é "claro que aqueles que estão no lado vitorioso estão jubilosos enquanto que aqueles no lado perdedor estão acabados". O argumento segue indicando que "ambos reagem ao mesmo evento ... é sua interpretação do evento que muda". Esta linha de raciocínio separa cuidadosamente os jogadores, torcedores e evento, pontualmente definidos em um instante em que a inferência lógica se realiza e "resultado", "júbilo", e "sentimento de derrota" são objetivamente produzidos e repartidos conforme a regra. Da cuidadosa purificação argumentada sobre o caso hipotético, longe está o processo, descrito em seus emocionantes detalhes em Damo (Damo, 2005, chap. 10), desenrolado na história de cada jogador e dos torcedores, que os traz para o acontecimento candente de sentir e expressar que é o jogo decisivo em um grande estádio – acontecimento durante o qual jogadores e os torcedores não são átomos

julgando cada momento por si, mas constituem uma amálgama de subjetividade e sensações compartilhadas e co-produzidas.

A *coerência* da teoria OCC também é resultado de uma estratégia cuidadosa: semelhantemente ao que vimos acima, as situações e seus julgamentos são argumentados sempre sobre casos hipotéticos. Ao sujeito real, que vive, não é dado espaço, e em seu lugar quem encena o teatro das emoções humanas é uma pessoa explicitamente genérica. Os casos argumentados são do tipo "Tome como exemplo uma mulher" (p.35), ou "Suponha uma pessoa" (44), ou "por exemplo, uma pessoa" (88), ou "uma secretária em uma grande multinacional" (p. 101). Interrogar este conjunto de casos hipotéticos produz um panorama esclarecedor do universo particular de situações que os autores, professores universitários em países desenvolvidos anglófonos, consideram relevantes e propiciadores de emoção. As situações trazidas por Ortony et al. dão conta de planos de vida e de cidadania legitimados e prestigiosos nesta sociedade, como "uma mulher que deseja tornar-se pianista", ou "uma pessoa que decidiu que deseja possuir, um dia, um Rolls-Royce ou tornar-se presidente dos Estados Unidos". Também presentes estão situações como "alguém descobrir que tem em haver um reembolso de U$100 de um órgão fiscal" (p.52), ou a secretária da multinacional cujo "presidente recebe um aumento de US$200.000" (p.101), ou a mulher que "poderia descobrir que ... para ganhar um carro ... deveria ter enviado pelo correio" um cupom de promoção (p.125).

Estes casos, tomados em conjunto, mostram uma configuração de vida e de valores que não são universais nem abstratos, e sim próximos do universo social dos cientistas que os expressam. Ao formular tipos de situações genéricas, e especificar através de exemplos que consideram paradigmáticos do que conta como atendendo o critério de pertencimento para estas situações, os autores também constroem um *vademecum* moral de função implicitamente pedagógica. Abrangente e completo, este guia moral elenca *quais* são situações que merecem sentimentos, e *como* sentir-se em relação a estas situações, até o ponto de invalidar a possibilidade de reações alternativas que não conformem com a regra (como é o caso da gratidão em relação a resultado não intencional, acima). Os autores, plenamente a par de que julgamentos estão intimamente relacionados a regras, ou padrões, apresentam padrões (morais) em alguns casos específicos, por exemplo quando mencionam algum "padrão abstrato como PESSOAS DEVEM REALIZAR PLENAMENTE SEU POTENCIAL ou PESSOAS DEVEM ESFORÇAR-

SE PELA EXCELÊNCIA". São abstratos na medida em que não especificam sua aplicabilidade, mas por outro lado são reais e existentes como padrões de normatividade moral familiares a nós, aos participantes, e aos pesquisadores autores do livro, sem gozar de uma universalidade *necessária*.

Os limites da exposição de regramentos morais abstratos, baseados em uma ética culturalmente localizada como fundamento da vivência psicológica, sem enriquecê-la de reflexividade e de perspectiva em relação a outras formas de *ser* o humano, revelam-se em momentos em que o fazer científico não dá conta de abranger com sensibilidade o viver *atualizado*, vivido, do humano. Isto parece ocorrer, por exemplo, no exame de situação em que, recorrendo a um sujeito novamente hipotético, Ortony et al. instam o leitor a imaginar uma situação de guerra, em que "alguém está deitado na cama ... ouvindo bombas explodindo à distância ... essa pessoa conclui, confiantemente, que sua casa ... não será atingida por estar distante dos alvos principais". Após estabelecer um contexto prévio, os autores prosseguem na narrativa exemplar: "Agora suponha que ela estava errada, e que sua casa é atingida". A reação emotiva, na visão da teoria OCC, é clara e purificada: "A mulher vai ficar surpresa". O que torna mais fácil purificar de maneira tão completa esta descrição, possivelmente, além da enunciação em modo hipotético, é a situação dos autores, em uma realidade distante do sofrimento e da perda de certezas e referências trazidas no rastro da guerra – sobre guerra e a implicação da distância e da proximidade de sua enunciação, especialmente pela imagem, ver (Sontag, 2003).

O campo de emoções e situações construído por Ortony et al. nos parece familiar, e esta é a dificuldade para entendê-lo como peculiar de um mundo específico. Os raciocínios apresentados e os sentimentos que são suas consequências têm, por isto, sua legitimidade. O argumento que nós trazemos não é o de invalidar a teoria e suas razões sobre o afeto; é, pelo contrário, afirmar sua legitimidade, mas uma legitimidade em um modo de ser diferente daquele que postula uma universalidade e uma objetividade totais do saber psicológico, e por consequência, do fazer da inteligência artificial. Compreender claramente esta verdade e esta legitimidade como situadas, válidas mas sujeitas a encontros com outras validades, é um passo necessário para respeitar e dialogar com outras culturas e outras formas de organização do social. Observar atentamente as situações propostas esclarece também nossa posição, como pesquisadores ligados à IA, enquanto próximos da construção de mundo da ciência objetiva e universalista e por este

motivo sujeitos a naturalizar e a tomar como universais esta construção. Repensar o modo de ser da validade e da legitimidade é importante também para que desencontros entre a teoria e o que inquirimos do mundo possam ocorrer, sem que isto coloque em questão uma verdade essencial da teoria... ou, igualmente, das respostas que recebemos do mundo. Retribuir ao mundo com respostas sensíveis, não apenas sensatas, também é fazer ciência e ser pesquisador.

### 7.1.2 Agentes artificiais e emoções

A emoção é referida no cotidiano do grupo como um fenômeno dado e estável, isto é, delimitado e conhecido, que é pensado em termos de sua influência ou consequência em atividades que deseja-se que sejam realizadas, e em termos de sua causação por eventos ou expectativas. Menciona-se, então, um estado emocional como um dado, a partir do qual decisões ou linhas de ação são tomadas: "o estado emocional... pra depois o agente tomar isso em conta...". Emoção, então, como estado interno de um ente, é um conceito particularmente apto a ser construído computacionalmente, incluindo a característica de modificar saídas em função deste estado.

Uma vez que a questão da emoção é estabelecida em termos de estado interno e causa de determinadas condutas computacionais, podemos observar como esta questão é detalhada e resolvida como problema computacional. Dentro do grupo, as soluções não são a mesma, nem homogêneas, mas são semelhantes e possuem pontos chave em comum.

A forma como o grupo entende e trabalha a emoção está muito vinculada à maneira como a questão foi tratada do ponto de vista da programação computacional. Alguns projetos anteriores, muito importantes na história do grupo, foram estabelecendo tanto um quadro de referência conceitual como um conjunto de implementações – código de computador e também modelos do problema – que hoje são utilizados e referidos com frequência. Estes trabalhos constroem um modelo para as emoções para ser usado pelos agentes criados, a partir de uma série de referências teóricas em várias áreas.

É saliente, ao observar-se esta forma de entender o fenômeno afetivo, a sua configuração em termos de estados – próprios do ente individual: o agente – e a causalidade prescrita para as transições entre estes estados. A emoção, a afetividade, não

são um acontecimento isolado na história do agente, já que outras características e condutas do agente são ditas relacionadas ou dependentes (da emoção); no entanto, estas relações ou dependências são, como visto, funções avaliadas. Emoção é resultado.

Para dar conta do desafio de realizar na prática agentes artificiais, implementando-os como programas de computador, as referências de modelo e teoria sobre o afeto são examinadas, e a solução encontrada passa por realizar uma seleção dentro do universo proposto na teoria. A seleção realizada é bem focada. Para realizar a seleção, um cenário aplicativo e, dentro deste, um cenário emocional são propostos; pensa-se então sobre como o agente deve se sentir frente às diversas situações do cenário. Os cenários costumam ser orientados a objetivos a serem alcançados, o que faz sentido dentro de um contexto em que o desenvolvimento de produtos de software é justificado por sua aplicação e utilidade.

O universo de emoções que é considerado, então, é elaborado em torno de um conjunto de "estados", pensados como a descrição adequada de um repertório em cada projeto de agentes. Isto é, a seleção realizada caracteriza-se por privilegiar sentimentos avaliativos – a serem esperados do agente artificial, vinculados aos entes presentes em sua representação computacional de mundo: outros agentes, a atividade em andamento em diferentes aspectos, o usuário. Os participantes, enquanto projetistas de sistemas computacionais, possuem uma noção precisa do que conta como diversidade emocional; são considerados como "sentires" para o agente os resultados relacionados a avaliações de regras do que o agente deve realizar, de acordo com seu modelo funcional ou a tarefa para a qual é orientado.

Agentes em jogos, por exemplo, modulam seu estado emocional em função do estado do jogo. Em um dos projetos acompanhados dentro do grupo, na verdade um conjunto de projetos relacionados entre si, um agente que joga xadrez estabelecia um estado interno baseado em uma função que calcula matematicamente uma "situação" do jogo – isto é, é uma função que retorna um valor que tem a intenção de indicar se o estado atual do jogo é favorável ou desfavorável ao contendor.

Além de jogos, outra plataforma para desenvolvimento de agentes artificiais, utilizada pelo grupo, são as animações, já mencionadas no capítulo anterior. Animações são clips de vídeo, de duração variável, mas em geral curtos, em que figuram cenografia e personagens desenhados por computador. O mecanismo e a estrutura utilizados para gerar

estas animações são semelhantes aos de jogos, no sentido de que há personagens cujos avatares realizam várias ações, incluindo interação com outros personagens, dentro de diferentes cenários artificialmente construídos. Há diferentes caracterizações para estas animações, desde aquelas com uma apresentação próxima do universo infantil de desenhos animados até outras semelhantes a jogos em ambientes que remetem à ficção científica. Há uma proximidade com o conceito de jogos; em vários casos, o mecanismo e a estrutura são construídos tendo em conta o propósito de serem apresentados a usuários, os quais interagem com o sistema e seus personagens como com um jogo de computador ou consola, através de teclado ou outros dispositivos de interação. Em outros casos o sistema é colocado em funcionamento, gerando uma sequência de vídeo que é gravada e depois examinada, apresentada a outras pessoas, ou divulgada. Este procedimento de gravar sequências também é utilizado durante o desenvolvimento dos sistemas-jogo, e para sua divulgação. As personagens presentes nestas animações são constituídas por uma apresentação gráfica e por um agente. A apresentação gráfica de personagens possui características de movimento para dar conta de apresentar as ações do personagem. Apresentações utilizados nas animações do grupo são de diversos graus de sofisticação; há desde simples desenhos de uma face sobre um fundo preto, em que traços da fisionomia como boca e olhos movem-se seguindo a fala do personagem, até representações mais complexas, em gráficos 3D, em que várias partes do corpo movem-se atendendo ações corporais como caminhar, mover mãos durante conversa e olhar para o lado.

Um sistema deste tipo foi desenvolvido como um software de geração de narrativa, para crianças em idade escolar. Uma história é contada a partir de uma animação com diversos personagens. O interessante é que a história não é pré-roteirizada; os eventos na animação vão acontecendo como reação a eventos anteriores, baseados em características dos personagens, e em intervenções da pessoa que está assistindo. A geração em tempo de execução, não prévia e não-roteirizada, de histórias animadas, através de inteligência artificial, é chamada de “narrativa emergente”.

A história, no caso deste software, gira em torno de uma turma de crianças na escola; há uma criança agressiva, que agride uma outra; uma terceira criança assiste ao acontecimento e tenta intervir. A cada personagem é associado um “plano”, um constructo que elenca e define uma sequência de ações que devem ser realizadas – como um guia para a conduta observável desta personagem. As condutas dos personagens da animação

vão sendo geradas como resultado do processamento de planos e avaliações do agente-personagem sobre os eventos como percebidos no modelo computacional. A discussão sobre emoções, então, centra-se em um processamento da relação entre eventos e planos, gerando uma avaliação que os relaciona – e a esta avaliação é dado um nome correspondente à lógica de "expectativa relevante" conforme o esquema teórico OCC (ver acima). Isto é, avaliando sobre um tempo futuro, esperança marca uma expectativa de algo desejado, que uma intenção seja realizada; medo marca uma expectativa de algo que se sabe que é ruim. A estrutura avaliativa, ainda seguindo o esquema OCC, também incide sobre a avaliação de objetivo realizado/não realizado, isto é, é dada uma avaliação ao passado. Os estados utilizados no caso deste trabalho são a expectativas do desejado, que confirmadas dão lugar a satisfação e não-confirmadas, a desapontamento; e a expectativa do indesejado, que confirmadas marcam receio-confirmado, e não-confirmadas marcam alívio (ver Quadro 3).

Emoção – ao menos os seis rótulos de emoção na rubrica de "prospecto relevante", que são o subconjunto considerado para este projeto – estão fortemente associados, portanto, a planos em existência, por parte do agente-personagem, constituindo marcadores da avaliação dos eventos relevantes para estes planejamentos. A proposição de um modelo funcional para o agente segue de perto a teoria OCC com bons motivos, uma vez que OCC se propõe como uma teoria cognitiva das emoções – articulando emoções em torno de valorações que o sujeito realiza sobre o que está em curso em seu universo. Um passo seguinte desta exploração, pelos pesquisadores neste projeto, foi apontar o que enxergam como uma lacuna na teoria OCC: "[a teoria] não *especifica* como emoções afetam raciocínio/cognição e seleção da ação" (ênfase adicionada); em outras palavras, a teoria não é fechada o suficiente, já que esgota a produção de emoções a partir da cognição, mas ainda deixa em aberto a especificação no sentido inverso. A solução foi encontrada, neste caso, em uma forma de colocar a emoção ainda mais fortemente vinculada a uma cadeia de causalidades/conseqüencialidades ligadas a plano e execução de objetivos. A partir de Sloman (1998), a emoção é formulada como um "mecanismo eficiente de controle". Sloman, no caminho complementar de Ortony e colegas, que propõe uma teoria psicológica que possa ser usada pela ciência da computação, trabalha a partir da inteligência artificial para propor uma teoria que seja aplicável aos domínios da psicologia e da biologia. Sua proposta é trazida sem dificuldade aparente para dar apoio à

modelagem do agente e a fechar o laço de funcionalidade para o estado emoção. O mecanismo de avaliação, ao operar sobre planos prospectivos, conceitualmente subsumidos a intenções, produz emoções de diversas intensidades, e aqueles que deram origem às maiores intensidades são selecionadas e prosseguem sendo processados. O fluxo de planos é coerente com a avaliação, formulada pelos desenvolvedores e programada no software, de seu desenvolvimento.

Em termos de narrativa presente nas animações, no desenrolar da ação que ocorre na animação com os agentes não há um roteiro fixo prévio; este mecanismo de processamento de planos do agente e a avaliação do estado em que o plano se encontra e dos eventos a que o agente tem acesso é que vão disparando condutas programadas. O cenário se desenvolve dentro de algumas restrições para o argumento da história: há um personagem agressor, que aqui será chamado Lucca, e há um personagem vítima, que receberá o nome de Giovanni. Têm nome (e, dentro do presente texto, pseudônimos), porque são *dramatis personae* de pleno direito, ficções verdadeiras. Planos disponíveis, estratégias de interpretação dos eventos, e regras de avaliação resultando em estados nominados com labels de emoção, para cada personagem, são construídos de maneira que ocorra um desenvolvimento da ação dentro do esquema previsto. O requisito, dado pela descrição prévia da animação, é uma história em que o personagem Lucca será agressor, e o personagem Giovanni será vítima. A narrativa é emergente, mas é delimitada por uma configuração que produza a história desejada.

É durante o decorrer da ação que um outro importante papel funcional dos estados nominados com emoções vai aparecer: a disponibilidade de figurações representativas nos traços de fisionomia, e das ações desempenhadas pelos personagens, dentro do vídeo de animação sendo gerado, como sinais exteriores visíveis para serem lidos como fenômenos afetivos e para comporem desta forma a narrativa. A narrativa é dada a ver, e podemos *ver* as emoções na figuração desenhada e nos atos dos personagens: a fala que descreve os atos, o empurrão, o desenho da boca com os cantos caídos. Entendemos a história, entendemos os personagens: a hábil construção visual da cena encontra a nossa familiaridade para, prazerosamente, ver e compreender desenhos animados.

Mesmo que não pré-roteirizada (além desta versão sem roteiro prévio, o software foi produzido também com uma versão pré-roteirizada), a história é simples, e suas variações

giram em torno de um personagem agressor e um que é agredido. Recapitula e reitera em seu decorrer uma dicotomia que tenta garantir a construção de uma história de acordo com o planejado. Expõe, além disso, uma perspectiva "planejada" sobre a compreensão do universo de conflitos entre pessoas, de crianças neste caso, e do assumir papéis no decorrer da vivência destas pessoas, no contexto de um mundo escolar bem organizado, cujas regras são visíveis e compreensíveis para os desenvolvedores e para os pesquisadores que orientaram a pesquisa de cunho psicológico envolvido.

A vinculação próxima entre sentimento e planejamento, colocada na conceituação teórica e na realização computacional de condutas para personagens, continua a desdobrar-se no olhar da pesquisa sobre a audiência – isto é, na linguagem da informática, os usuários – interessada que está em saber sobre a capacidade do software em contar histórias sobre o tema proposto. A história é construída usando, como artifício de projeto, trajetórias vinculadas aos agentes-personagens, as quais seguem duas vertentes de uma ravina, opostas, separadas mas unidas no seu vértice. Duas trajetórias que mantém-se em suas identidades, dadas e julgadas a priori, e não-intercambiáveis: o agressor e a vítima, o violento e o pacífico, a ameaça e o socialmente inapto. O software vai ser avaliado a partir de sua capacidade, enquanto artefato contador de histórias, de conduzir seus sujeitos-leitores-espectadores fielmente dentro destas trilhas-vertentes. Os pesquisadores, por fim, e de acordo com este desenvolvimento, expressam a expectativa que Lucca seja o mais detestado, e também que seja alvo de zanga; e que Giovanni, alvo de pena, seja o mais gostado.

Novamente, deve ser ressaltado que não colocamos em questão a legitimidade do sentimento, ou de zangar-se com um agressor, sentir pena da vítima ou de que sejam feitos juízos sobre a situação de agressão apresentada. O que está em destaque é a maneira como o campo de possibilidades foi estabelecido; a história, emergente e anunciada com as possibilidades de abertura a isto relacionadas, no entanto urge seu público a uma escolha, a um julgamento, esperado e conhecido. A escolha deste cenário não é casual, na medida em que responde a imperativos relacionados à pesquisa da área da psicologia que é realizada em conjunto com a pesquisa da ciência da computação, em inteligência artificial. Observa-se então que as escolhas que dão a fisionomia ao projeto encaixam-se harmoniosamente. Os pressupostos que motivam a busca sobre o afetivo, em inteligência artificial, são trazidos de uma tradição psicológica que funda sua compreensão do

fenômeno afetivo na avaliação cognitiva, e que reconhece seu objetivo de ordená-lo dentro de uma lógica da prática computacional. Estes pressupostos são trazidos, pelos praticantes da IA, para seu próprio contexto de pressupostos e requisitos, incluindo a formulação dos fenômenos humanos em termos de processamento de informação, e a percepção de que é necessário um modelo claramente definido para ser implementado. O encontro destas propostas dá forma ao sistema, em que as posições dos personagens são clivadas pelo binário bom/mau, desejável/indesejável. E, por fim, no encontro do artefato com as pessoas, as expectativas que eram colocadas sobre a máquina/roteiro/agente são reiteradas sobre as pessoas. Há duas expectativas sobre o sentimento das pessoas: uma de que aconteça a reação dupla pena/ira, a outra é de que sejam direcionados sobre determinados personagens. As pessoas, agora, serão solicitadas, perante a história emergente, a realizar uma escolha, já aguardada e julgada, dentro dos confins de um espaço emocional previamente organizado pelos proponentes do sistema.

A atribuição prévia de um "script afetivo" a personagens mescla-se com atribuição de scripts a pessoas. Há uma produção hábil e engenhosa das personagens para que sejam inteligíveis e para que passem um sentimento de verdade (ou plausibilidade): "quando a vítima está muito triste ela tende a chorar, enquanto que quando o agressor expressará sua tristeza de uma maneira completamente diferente" (de um trabalho do grupo). Na personagem está a hábil produção e demonstração, para que possamos melhor e mais facilmente ler a personagem. No entanto, a construção, segundo a IA, prossegue no raciocínio de que seleciona-se uma gama de estados internos atribuídos ao agente e enunciados na animação. Em um passo contínuo, e partindo da premissa de que emoções e sentimentos *são* estados internos, discretos e distintos, imputa-se ao leitor/espectador, que interage, a mesma gama e natureza de estados internos, apagando a hábil e engenhos *leitura* e interpretação feitas pelo usuário sobre a animação. Uma das consequências deste processo é que, quando a leitura da pessoa, enquanto espectador ou usuário, não fecha com a expectativa que os pesquisadores possuem, isto é considerado erro ou falha, e que deve ser buscado no usuário ou no projeto do sistema. A produção subjetiva fora da norma esperada, deste ponto de vista, não é hábil ou engenhosa, e sim um problema a ser controlado.

## *7.2 Rituais e sociabilidades*

> Ritual do Jantar – este ritual é ativado após todos os convidados chegarem à festa. Consiste no anúncio, do anfitrião aos personagens, de que o jantar iniciará e que todos devem tomar seus assentos. Então o ritual continua com os personagens sentando à mesa e começando a comer. Entretanto, enquanto que em uma cultura todos acorrem à mesa imediatamente, nem mesmo esperando pelo anfitrião terminar o anúncio, na outra cultura todos devem esperar antes que o ancião tome seu lugar antes que possam sentar-se, e então devem esperar o ancião começar a comer antes que possam comer. Além disso, o ancião nesta cultura tem o privilégio de sentar-se na cadeira adornada. (de um dos trabalhos do grupo)

Assim é descrito, em um trabalho escrito da produção do grupo, um "ritual sintético". De maneira semelhante às performances de agentes artificiais já descritas, trata-se da realização de uma animação gráfica, isto é, visualmente apresentada em um vídeo, em que agentes-personagens desempenham uma série de ações. As animações a qui discutidas são, novamente, do tipo descrito pelos participantes como "emergentes", isto é, não têm roteiro fixado previamente; em vez disso, resultam da programação de planos para cada agente, e da interação em tempo de execução dos diversos planos dos agentes (em outras palavras, "emergente" aqui significa uma atualização não roteirizada previamente, a partir da interação das possibilidades do sistema).

A cultura, como realização da distintividade e da diversidade humana, chama a atenção da IA como traço do qual dar conta em termos de replicação tecnológica. O grupo apresenta um conjunto de trabalhos em que são explorados o tema da cultura e as maneiras como esta é capaz de dar forma ao espaço de ação do humano. A motivação é justamente a percepção da cultura como caracterizador do ser humano, como provocador de diferenças. Chegando, como tema a ser explorado na IA, depois da lógica, da cognição que processando informação, e da emoção, é colocado como em um "anel externo" de distinção. Ou seja, de uma certa forma cultura é externa na caracterização do humano, compondo conjuntos de regras – externas – a partir das quais os processos de cognição e emoção, mais internos, serão disparados.

A exploração proposta movimenta-se em um sentido de construir agentes-personagens que expressem cultura. Estabelecer o ponto de partida é a dificuldade inicial

para um empreendimento como este, e por este motivo é interessante prestar atenção a ele. Cultura, por ser um tema amplo e não fechado, requer tomar algumas premissas de partida para a partir delas, disparar propostas de como trabalhar.

Há um ponto de partida visível neste esforço, que é, como anteriormente visto com o universo afetivo, a tensão peculiar que há entre o projeto de criar agentes-personagens artificiais e o humano que é seu modelo e ao mesmo tempo sua audiência. O afã de preencher, de completar estes agentes de maneira que possam ser sujeitos das várias relações com humanos que se tencionam – que sejam reconhecíveis, plausíveis ("conseguir ter comportamento credível"), que possam ser utilizados como suporte de performances pedagógicas e lúdicas diversas – leva ao questionamento, centrado no agente, de qual elenco de performances deve ser capaz, visível no ato de explorar a ampliação deste elenco:

> "até agora nós nos tínhamos preocupado imenso com a personalidade, com as emoções, e como é que isso afetava o comportamento, tínhamos utilizado estudos sobre a personalidade, da parte de psicologia sobre a emoção [...] mas a ideia da cultura ainda não tinha sido explorada, e a cultura também é uma parte muito importante da forma como influencia nosso comportamento" (participante deste projeto)

A cultura, aqui, entra como um modificador, um causador de condutas, e que por isso deve ser levado em conta, isto é, programado computacionalmente. A importância de abordar – e ser capaz de dar conta de – este traço é reforçada com justificativas de aplicação prática, tais como "aplicações para treinamento intercultural".

Uma estratégia de caracterização substantiva para constituir uma cultura implementável em agentes artificiais articula-se, neste trabalho do grupo, com o quadro conceitual que foi desenvolvido por G. Hofstede a partir dos anos 1970 (Hofstede, 1980). Este autor, pesquisador holandês, propôs uma teoria da cultura baseada em dimensões culturais que a caracterizariam. Estas dimensões, atributos mensuráveis (comparáveis) de uma cultura a outra, seriam 5: distância de poder, individualismo/coletivismo, masculinidade/feminilidade, evitamento da incerteza, e orientação a curto prazo ou longo prazo. A cultura, nesta perspectiva, colore o sujeito que a ela pertence, através de certos itens – elementos – distintos: "elements that characterize an agent of a certain culture", nas palavras de um texto do grupo. Os elementos desta cultura são substantivos e

enumeráveis, mesmo que não "materiais", como rituais, símbolos, e a avaliação de situações. Para dar forma substantiva à cultura foi escolhido como objeto o ritual: rituais que possam obter comportamento cultural reconhecível para seus agentes (grupos de agentes) artificiais.

Escolhido como elemento visível constituinte da cultura, enunciável em separado, o ritual vai ser pensado como um índice pelo qual a cultura – específica e delimitada – do agente-personagem possa ser dada a reconhecer; dentre "elementos que caracterizam um agente de uma cultura [...] vamos focar [...] rituais". Escolhido, por este motivo, como objeto a ser implementado, merece também uma definição conceitual – embora os proponentes do trabalho reconheçam que este é um conceito que resiste a ser definido da maneira como se deseja. Para esclarecer do que se deseja falar ao mencionar "rituais", descrições são dadas de maneira a que, paulatinamente, seja possível compreender que tipo de fenômeno humano se trata. Rituais, então, são descritos como atividades; compostos de ações; ou como conjunto de atividades. As atividades relacionadas a rituais são definidas a partir de uma distinção sobre seu caráter ou finalidade, como um contraponto a atividades não-rituais, que são utilitárias ou técnicas. Atividades técnicas são espontâneas; atividades rituais, rotinizadas e regradas; as técnicas são eficazes, as rituais simbólicas e não-instrumentais. Apesar de todas estas diferenças entre atividades técnicas e rituais, a proposta para implementar rituais computacionalmente, neste trabalho, retorna ao conceito, próprio da inteligência artificial, de *planos*. A descrição de planos como sequência de passos a serem executados revela, aos olhos do projetista, a semelhança com a descrição do ritual como conjunto de ações marcados como receita precisa de atividades. A semelhança entre ambos, segundo esta noção, deve ser temperada por uma diferença, já que a atividade técnica não parece com a atividade ritual, ou não deve parecer. Rituais, portanto, são diferentes porque são baseados em *atividades rituais*, e a distinção que pode ser apontada é a importância dada à sequência de passos (isto é, a liturgia, em oposição aos planos tradicionais, em que o importante é o resultado final. A implementação tecnológica do ritual foi proposta, dada a caracterização construída para o ritual, como uma adaptação ao conceito de plano oriundo da inteligência artificial para dar conta da execução de um conjunto de atividades da maneira caracterizada.

A forma dada ao projeto de tratar a cultura a partir da IA, conforme apresentada, demonstra uma busca de demarcação e categorização prévia. O motivo para esta atitude

perante o problema tem sua raiz no próprio movimento com que o tema foi abordado: há um fenômeno, o ritual, que caracteriza o humano, e que é importante para sua sociabilidade, e por isso deseja-se construir uma sua réplica de maneira programática, computacional. Dado isto, faz parte do projeto demarcar e enumerar as características suficientes para que se possa planejar uma animação de agentes que, ao ser apresentada, possa ser lida e identificada, sem maiores dúvidas, como este fenômeno.

Construir o fenômeno selecionado de maneira computacional através de agentes artificiais constituía-se, no quadro temporal em que a pesquisa era realizada, um feito distintivamente inovador. De fato, esta foi uma motivação citada explicitamente para realizar a pesquisa: a possibilidade de inovar, valor precioso no campo tecnológico. Como consequência, teorias e referenciais que pudessem dar apoio a este projeto, dentro das disciplinas de ciências sociais e humanas como psicologia e antropologia, não haviam sido previamente mapeada, nem – à parte os estudos de Hofstede – havia uma tradição ou referências-chave já adequadamente apropriadas pela computação, como era o caso da computação afetiva, conforme apresentado na seção anterior. Isto torna ainda mais interessante a forma como a literatura sobre o tema, por exemplo (C. M. Bell, 1997), foi mapeada, dando forma, com cuidado, a um cenário que servisse não apenas como uma descrição (posterior ) do ente "ritual", mas principalmente como um guia (prévio ) gerador de planos que coubessem na descrição. Daí se destaca o conjunto de oposições com potencial de caracterizar convenientemente a atividade que se desejava planejar.

A realização obtida é o "ritual sintético", ou seja, a animação com os agentes-personagens realizando uma ação concertada que deve ser lida e interpretada, como culturalmente distintiva, pelo espectador. O caráter coletivo da ação, e, conforme programado, as peculiaridades diferencialmente apresentadas, dão conta de expressar algo sobre a ação que vai além do "individual" de cada agente-personagem. No entanto, a descrição pela qual os próprios participantes procuram dar conta da ação programada, e que é programada para *expressar* algo, pode ser examinada em contraste com a conceitualização construída, a fim de tornar visíveis algumas dificuldades do processo adotado de conceitualização instrumental, adotado e justificado pelos participantes. Os remédios adotados pelos pesquisadores para enriquecer e garantir o sucesso da animação, expressiva enquanto "ritual e cultura", são uma das pistas que podem ser seguidas no presente caso.

Rituais, no contexto proposto, são progressivamente delimitados no procedimento definidor adotado, mas ao serem definidos não cedem uma diferença clara em relação a outras atividades. Para que possam ser computacionalmente construídos – programados previamente de acordo com certas intenções expressivas dos participantes – julgou-se necessário insistir em uma definição externa, enunciável e repetível independentemente do observador (independente do sujeito, focada no objeto); é dessa forma que são lidas as contribuições da literatura antropológica, ao trazer, por exemplo, para o rol definidor do fenômeno a "invariância" de que trata Catherine Bell (C. M. Bell, 1997) em relação a eventos com características rituais. O resultado desta objetividade é que, afinal, para a IA atividades rituais são como outras atividades que seguem um plano individual pré-definido (com a peculiar translação teleológica relacionada à proeminência da "liturgia" e não da "finalidade"). Em um contraste, as atividades programadas na animação, descritas no Ritual do Jantar, são atividades que, embora cotidianas, são prontamente associadas a diferenças em atitude do coletivo, isto é, relacionadas a cultura, e não a particularidades individuais. Saudações e maneiras à mesa, ao serem destacadas, expressam o que os pesquisadores desejam expressar, isto é, diferenças de cultura, ou a própria cultura enquanto chave de distintividade.

Uma mudança de perspectiva, assim como a empreendida no caso da emoção computacional, pode auxiliar a desatar o busílis colocado pela definição apriorística rigorosa que foi empreendida. O ponto estratégico para esta perspectiva pode ser encontrada na própria maneira como os participantes veem o acontecer da ação na animação: a ação é uma *narrativa* realizada por *personagens*, que possuem *papéis*. Se, colocada desta forma, a questão do ritual parece remeter a uma performance, a um desempenho dramático, é exatamente esta uma das formas que a antropologia propõe para compreender este fenômeno social (Peirano, 2000; Tambiah, 1985). Mariza Peirano (Peirano, 2000) retoma, a partir de Tambiah e Turner, esta forma de abordar a questão, ao propor considerar a diferença entre a ação cotidiana e a ação ritual como a coloca o nativo, isto é, a própria pessoa que está a engajar-se em uma ação diferenciada e mais formalizada, e que a sente como especial. Então, passa-se a observar estes eventos a partir do recorte realizado em termos nativos, percebendo como são ordenados, organizados, e como ganham "um sentido de acontecimento cujo propósito é coletivo" (Peirano, 2006, p. 130). Este sentido coletivo está imerso nas premissas culturais e nas sínteses, elaboradas a

partir da visão de mundo, partilhadas, permitindo entrever – embora não de maneira sempre explícita, e não meramente referencial – categorias cosmológicas e sociais desta cultura. Não é o caso que a cultura simplesmente explique o ritual, ou que certa crença determine um rito; mas que "cosmological constructs are embedded [...] in rites, and that rites in turn enact and incarnate cosmological conceptions" (TAMBIAH, 1985).

Rituais, desta maneira, podem ser entendidos como "sistemas culturalmente construídos de comunicação simbólica", "bons para pensar e bons para agir", (PEIRANO, 2006) e não apenas reflexo ou decalque cultural. A performance constitui o evento ritual entrelaçando a forma e o conteúdo, utilizando múltiplas maneiras de expressão e levando o participante a experimentá-lo de maneira intensa; o ritual, segundo Tambiah, é performativo ainda em um outro sentido: em que certos casos dizer é realizar um ato e não apenas comunicar, e no sentido de remeter a valores indexicalmente percebidos pelos participantes (TAMBIAH, 1985).

Retornando ao tema do ritual sintético, e retomando a partir destas considerações, é possível analisar tanto o percurso pelo qual foi buscada uma definição generativamente satisfatória para a performance emergente dos personagens sintéticos, como a solução interessante e ad hoc adotada. Rituais são distintos e distinguíveis por serem especiais, e por serem vividos desta forma, pelos participantes do evento. Rituais não são apenas simbólicos, porque não estão com a cultura em uma relação de expressão ou significação (somente), mas em uma série de relações mais abertas, tais relações de índice (marca visível de algo que há), alegoria (dizer de algo além), de performance (fazer acontecer, fazer ser), de reiteração, de pertencimento e de metonímia. O ponto importante desta análise é que, ao procurar construir uma definição que não apenas reconheça, mas que possa ser instrumental para gerar o evento, e que ao mesmo tempo seja objetiva, isto é, focada no objeto (ritual) e não dependa do sujeito, uma realização deste tipo corre o risco de excluir o íntimo porquê do ritual, enunciado nas pessoas que o realizam e no seu vivenciar do evento.

Para preencher esta espécie de vazio de sujeito enunciador deixado pelo processo de definição percorrido, foi construída uma performance que remetesse a um senso coletivo, organizado de maneira que seu decorrer fosse moldado por diferentes formas como os atores-personagens-agentes veem o outro presente em relação a si e em relação ao

coletivo. Esta performance coloca-se em um campo de tensões interpretativas no mínimo curioso. A questão, que poderia ser colocada na linha de objetividade privilegiada pelos participantes, sobre se a performance é propriamente "ritual" é respondida, pela abordagem que propomos acima, pela colocação de que o recorte, a distinção como evento distinto foi colocada pelos próprios participantes-programadores. Encaixada a esta resposta, fica a pergunta: se é de fato significativa para os participantes-programadores, em que medida é uma performance relacionada a seu próprio universo, embora não reconhecido como tal? Ações partilhadas em coletivo do porte de rituais ganham sua intensidade através do viver-no-instante ritual dos participantes; a performance como ato cultural é eficaz na medida em que envolve, persuade e legitima o *ser cultural* do participante, mesmo que seja frente ao *outro cultural*. Há, então, um problema relacionado a quê, exatamente, é interpretado ou interpretável quando alguém está diante de uma animação com uma sequência de gestos que remete ao ritual, mas que não é animado, de maneira densa, pelo apelo performático de participantes. Há, parece, uma espécie de loop, no qual a performance, que nasce da intenção de dar a ver o culturalmente diferente, é encenada e animada – na verdade, meta-encenada por agentes-personagens na animação – a partir da vivência cultural dos participantes-programadores, sem que pessoas que encarnam algo diferente tenham participado no processo.

Esta questão tem desdobramentos interessantes, que podem ser exploradas considerando em conjunto a produção realizada em um projeto relacionado (eCircus Consortium, 2009). O projeto procura auxiliar na integração cultural de imigrantes e refugiados através de um jogo que, ao apresentar para o jogador uma cultura sintética, promova a empatia intercultural. Há disponível, na página do projeto, um curto vídeo promocional sobre o jogo (Latour, 1994), protagonizado, em um clima de ficção científica, por habitantes de um planeta distante, semelhantes a lagartos, com os quais o jogador deve interagir. Aqui, o enfoque não é sobre o ritual, mas sobre a expressão da cultura do outro – e, portanto, embora de maneira não explicitada ou reconhecida, sobre a interpretação da cultura do outro. Em ambos os casos, em função do objetivo destacado para cada projeto, a performance a ser animada pelos agentes-personagens é construída a partir da proposição de ações que atendam ao objetivo, isto é, ações que segundo os participantes-programadores sejam culturalmente específicas, contribuindo para tornar visíveis diferenças entre culturas. A performance é explicitamente colocada como a

metáfora pedagógica de um outro a ser conhecido e respeitado. O que ressalta, nesta proposta, é a elipse sofrida pela possibilidade de uma auto percepção do ponto de vista cultural: a diferença é encenada a partir do exótico presente no outro, preservando da exoticidade o contexto cultural do qual se parte, isto é, o dos participantes-programadores. Tudo se passa como se o laboratório de criação em inteligência artificial, com seus ocupantes cientistas, fosse uma "sala limpa"[12] capaz de referir-se, de maneira objetiva porque não culturalizada, a caracteres culturalmente impregnados, presentes do "lado de fora". Para o jogador-espectador-participante, a percepção e o conhecimento que ocorrem na interação apontam sempre para um outro estranho, remoto e plasmado a partir de itens de exótico: os gestos dos habitantes do planeta são apresentados como "strange gestures", seus costumes como "strange customs", e até mesmo a apreciação que o personagem faz do jogador-espectador coloca isto em pauta: "your face pleases me strangely".

A ênfase colocada é claramente em catalogar e tornar visíveis elementos discretos que possam servir como marcadores do cultural: "conjunto de comportamentos culturalmente específicos", "símbolos culturalmente específicos", "capturar especificidades da comunicação em uma cultura". A apresentação da performance dos agentes-personagens é desenvolvida com grande perícia técnica pelos autores, resultando em uma narrativa visual – tal como a do vídeo promocional – que é claramente inteligível e interessante, encaixando-se dentro dos paradigmas da narrativa em ficção científica tradicional, a qual dedica-se, entre outros temas, ao encontro com culturas "alienígenas". Ao enfatizar elementos explícitos e apresentá-los em uma narrativa atraente, no entanto, o resultado é que a culturalidade própria do artefato, e de seus criadores, é apagada sistematicamente. No universo destes projetos, traços interpretáveis como culturais sempre apontam para fora, para quem é alvo da descrição, e constituem como cultura reconhecível algo sempre fora do círculo do artefato-programa e dos seus cientistas-programadores – que, neste movimento, são objetivados, e tornados subjetivamente neutros e não-parciais. Esta estratégia de purificação (Latour, 1994) e de apagamento (Star

12 Sala especial para fabricação de chips eletrônicos, com sistemas especiais de depuração e filtragem que tornam o ambiente extremamente isento de partículas e impurezas. Por extensão, a programação de *software* sem que programadores entrem em contato com instâncias concomitantes da mesma funcionalidade.

& Strauss, 1999) de sua especificidade cultural normaliza ao extremo e faz invisíveis as práticas culturais dos participantes para eles mesmos.

Um caso notável para comparação são as reuniões de grupo, que são desenvolvidas todas as semanas, e da qual participam a professora líder, os professores participantes e os alunos e pesquisadores mais graduados (como por exemplo, os doutorandos). A reunião desenvolve-se segundo uma liturgia regular, reiterada a cada reunião, vivida e encenada com naturalidade por todos os participantes. Alguns minutos antes do horário da reunião, os alunos e pesquisadores começam a sair de suas salas de trabalho, e a procurar seus colegas em suas respectivas salas, convidando-os a passar de sua atividade cotidiana para um momento de expectativa, no corredor. Este convite costuma vir através de uma pergunta: "Vais à reunião?". Quando um grupo de colegas em espera se forma, por vezes junto com um ou outro professor, ocorre um deslocamento mais ou menos coletivo até a sala em que são realizadas as reuniões. Nesta sala, há um conjunto de lugares (mesa e cadeira de aula), dispostos em U, e um quadro na parede próxima à extremidade aberta do U. À medida que vão chegando, os alunos e professores vão tomando lugar, primeiro nas laterais da sala. Professores costumam completar seu número depois dos alunos, tendendo a ocupar os lugares do centro. Um lugar, em algum dos assentos próximos ao centro do U, é sempre deixado vago. Uma pessoa posta-se junto ao quadro, e com auxílio de algu ém presente, ajusta o seu computador a um projetor: é a pessoa que vai realizar a apresentação técnico-científica do dia. Conversas várias, comentando temas recentes ou antecipando temas da reunião, acontecem entrecruzadas, em um compasso de espera. Por fim, dez ou quinze minutos após o horário atribuído para a reunião, a professora líder chega, ocupa o lugar vago, próximo ao centro do U. Ao chegar, realiza um comentário rápido ao qual todos prestam atenção, e então ela, ou o aluno designado para gerenciar a agenda das reuniões, dá um ligeiro sinal em direção a quem vai apresentar. E, com um pequeno aceno que passa recibo do sinal, o apresentador inicia sua fala.

Com algumas variações contingenciais, este acontecimento repete-se quase todas as semanas. É um momento importante para o conjunto de pessoas e para seu funcionamento como grupo, como equipe. A importância pode ser medida pelo esforço dos participantes em comparecer, pelo esforço dedicado à preparação da apresentação pelo palestrante da vez, e por menções explícitas de participantes a esta importância. No entanto, apesar desta relevância, e da estrutura estável da performance do grupo, as

reuniões são descritas a partir de uma tipologia da naturalidade, da normalidade não saliente, e da instrumentalidade racional. A descrição dada desta atividade contrasta com a do "ritual do jantar" não pela minúcia de detalhes levados em conta, mas pela interpretação dada ao momento – a atribuição, ou não, de caráter ritual.

Cultura, portanto, apresenta-se para os participantes do campo através de elementos de peculiaridade cultural, tais como os rituais, que são atribuídos e considerados visíveis em *outros*, em contraponto a uma posição de certa forma "neutra", a posição da ciência e do artefato tecnológico que é tela em branco onde a peculiaridade pode ser projetada e conhecida. Esta maneira de descrever a culturalidade, que designa e distribui em polos binários a proximidade e a distância, o natural e o cultural, o normal e o exótico, o racional e o tradicional, o utilitário e o simbólico, remete ao orientalismo como formação discursiva estudada por Edward Saïd (MARCUS, 2007; SAÏD, 1990).

Segundo Saïd, o *orientalismo* constitui-se em um complexo e bem construído campo de saber textual sobre o Oriente, compreendendo desde textos acadêmicos até relatos de viagem, em que noções pré-configuradas sobre o Oriente como locus da diferença exemplar em relação a um Ocidente implícito são reiteradas: o Oriente é misterioso, irracional, caótico, entre outros atributos. O orientalismo surge como uma "constelação regular de ideias", sedimentado na produção de textos europeus sobre o Oriente a partir do século XVIII, em que a crescente produção de conhecimento vai sendo amarrada a noções "arquetípicas" sobre as pessoas, a sociedade e o território oriental. O orientalismo não é uma criação de ideias sobre o Oriente "sem realidade correspondente"; é, antes, a cristalização abrangente de uma forma de observar e interpretar este Oriente como um outro, como um conspícuo diferente. Saïd examina, através de um número de casos, como cada nova produção de texto sobre o tema oriental pressupõe a formação do conhecimento do autor deste texto dentro do campo, e ao mesmo tempo sedimenta e autoriza com força renovada este campo.

A análise de Saïd mostra como a produção científica articulou-se com a ocupação colonial ao longo dos séculos XIX e XX, ambas emprestando-se mutuamente autoridade para falar e agir sobre seu sujeito de conhecimento e poder, deixando muito pouco espaço para que as próprias pessoas sobre quem se falava pudessem colocar sua voz. Neste percurso, o Oriente foi sempre alvo, destino, objetivo, deste falar e agir, e sempre foi

mostrado como correspondendo a uma série de características culturais. A contrapartida "ocidental" frequentemente não é explicitada, constituindo-se em um binário oculto na relação com as qualidades atribuídas ao oriental. É assim que que se estabelece o "distante" Oriente como o lugar do misterioso, do tradicional, e por fim, daquilo sobre o que o especialista (ocidental) tem autoridade para dizer algo.

A distribuição realizada de agências e vozes (quem faz e é responsável por que ações, e quem fala o quê sobre quem), dentro dos projetos abordados relacionados a cultura sintética, traz à baila muitas das questões de Saïd. O fato de que o projeto é de âmbito europeu, tem o o objetivo de auxiliar na integração de refugiados, e tem por nome "Orient" é, em si, significativo. A produção eficaz de conhecimento e tecnologia provê ao mesmo tempo uma direção estabelecida para o fluxo de apresentação e enunciação. Há um grupo que inscreve características que (este grupo) vê como culturalmente específicas em artefatos tecnologicamente eficazes. Em outras palavras, cientistas da inteligência artificial estabelecem representações de "rituais" e outros "comportamentos culturalmente específicos" e os inscrevem em programas que geram narrativas visuais em jogos para computador. O jogo como consequência projeta uma arena em que se materializam e visibilizam as concepções destes cientistas sobre o que é considerado culturalmente específico e reconhecível – como cabanas com telhado de palha –, mas não é exposta a culturalidade do universo que cerca esta arena. A reunião de jovens para divertir-se com o jogo, mostrada no vídeo, o próprio processo de construir o jogo e o vídeo, e o cotidiano dos cientistas que os constroem não são atribuídos a nenhuma culturalidade "específica". Aparentemente, este é um cotidiano inespecífico, natural, "normal". E, fechando o ciclo de autoridade representacional, não há outras fontes de representação para o que constitui um "comportamento cultural", não há avenidas para o "outro" descreva a própria culturalidade dos pesquisadores, feita desta forma invisível e não-analisável.

## *7.3 De agentes, ambiguidades e traduções*

Para o fechamento deste capítulo, gostaríamos de trazer para a discussão a noção de *versões*, de Vinciane Despret (Despret, 2004). Como vimos, construir conceitos sobre o humano e afirmar sua existência de maneira objetiva, pura, traz algumas dificuldades. A emoção separada do sujeito, o ritual sem pessoas, o sentir positivo/negativo funcionam bem se indagados e protegidos por premissas e infraestruturas bem determinadas. A questão que colocamos é: neste contexto, são os sentimentos, ações, e significados *de quem* que estão ligados a estes conceitos? Enquanto a regra de formação da emoção é familiar, e propõe sentirmo-nos felizes por alcançar objetivos que, entre nós pesquisadores, aceitamos, tudo parece muito natural: ira *é* negativa, *queremos* ser excelentes, aquilo que amamos nos parece *interessante*. Da mesma forma, cabanas de palha e linguagens incompreensíveis são *típicas* de alguém distante, e esperamos que todos reconheçam a *vítima* como tal, e que sentimentos devemos ter por ela. Casos em que isto não ocorre são vistos, segue-se, como inválidos, *falsos* ou pertencentes a um *outro*, possivelmente *desfavorecido* ou *primitivo*.

Despret, propondo uma maneira diferente de compreender esta paisagem do humano, nos leva a perceber que haverá dificuldades na trajetória do conhecer a emoção, mas que não precisam necessariamente ser "vencidas"; compreender o diferente é uma questão de "tradução", com a responsabilidade de fazer-se entender em âmbitos diferentes que não são incomensuráveis, mas tampouco cedem a uma medida objetiva fora do sujeito, porque o *sujeito pesquisador* não se apaga simplesmente. Há o risco na tradução, que é desafio, e não solução assegurada.

Colocar em cena a tradução, e admitir outras versões do humano, abre espaços para compreender não somente porque outros respondem diferentemente, mas também porque respondemos de nossa forma. Despret mostra como experimentos psicológicos ao longo do tempo produzem objetividades sobre sujeitos trabalhando sobre o campo disponível para resposta (Despret, 2004, chap. 4); o que se produz são verdades que respondem a injunções feitas pelo pesquisador no peculiar ambiente social que é o laboratório – ou o questionário, que também encena um ambiente. Resultados de experimentos, portanto, não são falsos, mas têm uma localização, uma situação própria. Pessoas não são, neste

ponto de vista, apenas entes autônomos, atômicos e que seguem racionalidades, ou irracionalidades, individuais e privadas, ou alternativamente, externamente determinadas. Mesmo que isto constitua-se em um desafio para o trabalho científico e tecnológico, a experiência do sujeito é sempre plena das ambiguidades, hesitações e contramarchas tão próprias da vivência humana. Não dar espaço para versões *subjetivas*, isto é, originadas no próprio sujeito que se estuda, deixa apenas *um* caminho à compreensão que se deseja, se é desejada de fato, deste sujeito. Como exemplo, destacamos, em Ortony et al. (1990, p. 62) uma passagem em que os autores examinam um experimento realizado na década de 1960, no qual pessoas eram condicionadas a levar um choque ao ouvir um determinado som. As pessoas assustavam-se, e continuavam a mostrar-se assustadas ao ouvir este som mesmo depois de terem os eletrodos removidos e serem comunicadas de que não seriam administrados mais choques. Concluem os autores sobre esta continuação da apreensão que isto é "consequência emocional da *inabilidade* das pessoas em aceitarem situações como reais, mesmo em presença de evidência incontrovertível", e que "o senso de realidade gerado pelo medo era *inapropriadamente* alto" (ênfase minha). O sentir do sujeito, enunciado em seu corpo, vivido como historicidade dentro de um tempo no laboratório, não fecha com a regra e portanto é considerado inábil e inapropriado: há, aqui, um conflito de versões, e talvez falte um desejo da tradução.

Concluindo, o papel desta tradução não é converter o estranho, o distante, em denominador comum, e sim compreender, encontrar, aprender com o outro, ao invés de interpretá-lo, e utilizar das mesmas práticas para compreender-nos reflexivamente. Dentro desta prática, deixamos, os pesquisadores, de ser a medida, e passamos à responsabilidade de realizar o diálogo entre as várias formas de ser e sentir – ambíguas ou contraditórias em seu próprio direito. Trago aqui o exemplo em (Abu-Lughod, 1999), que estudou o gênero poético *ghinnāwa* entre os beduínos do deserto egípcio. Em uma sociedade regida por um código comportamental estrito, o da *modéstia*, sentimentos são vinculados à honra do grupo e à autoridade paterna e masculina. Mulheres usam o véu como expressão desta modéstia, e o discurso cotidiano expressa distância emocional entre homens e mulheres. No entanto, corre em paralelo uma forma de expressão, tradicional na forma e no conteúdo enquanto que reconstruído criativamente por seus praticantes, que é a poesia oral *ghinnāwa*. Este gênero é principalmente praticado pelas mulheres, e consiste em canções pungentes sobre relações amorosas, frequentemente trágicas, e muitas vezes

com um sentido alegórico, não abertamente reconhecido, em relação à vida e sentimentos de quem a entoa. Esta poesia ocupa um lugar ambivalente entre os beduínos, pois seus temas e abordagem entram em choque com a modéstia, virtude exemplar; *ghinnāwa* é respeitada e apreciada, e ao mesmo tempo considerada inadequada e mesmo subversiva (por exemplo, mulheres normalmente não a entoam em presença masculina). Quando Abu-Lughod voltava de seu trabalho de campo em uma longa viagem de carro pelo deserto, seu anfitrião, patriarca da família, entregou-lhe uma fita cassete e pediu que ela a ouvisse. Era uma poesia, a história verdadeira de um jovem separado de sua amada por um casamento arranjado seguindo interesses familiares. O jovem gravou o poema, e o enviou à amada (esta mesma cassete que Abu-Lughod estava ouvindo), que após ouvir a canção morreu depois de alguns dias, assim narrou o patriarca.

Aqui entrecruzam-se as questões que procuramos provocar: Abu-Lughod aprendeu com as mulheres sobre o sentimento velado de paixão, expresso na poesia, e o patriarca, responsável pela ordem moral incorporada pela modéstia, apresenta a ela, uma mulher, e estrangeira, uma instância poética que desafia esta mesma ordem. Uma translação de versões, e não o desejo da interpretação precisa, foi construída como compartilhamento situado da ordem de emoções em cena, permitindo aos envolvidos o diálogo significativo. Sentimentos vinculados a valores em relação múltipla, de coerências móveis, foram mutuamente oferecidos e compreendidos, em uma viagem memorável para Abu-Lughod, e construtiva para nós, pesquisadores da Inteligência Artificial.

# 8 Conclusão

Esta tese examinou uma série de noções e premissas sobre o humano presentes no trabalho da Inteligência Artificial, através de um conjunto de casos de construção de sistemas computacionais inteligentes, ao redor das atividades de um grupo de pesquisa em Agentes Inteligentes. Noções examinadas, explicitamente apontadas dentro do grupo como objetivos relevantes para seus sistemas, foram o conhecimento de senso comum, a emoção, a cultura, a agência e o encontro de agentes com usuários.

Dentro da Inteligência Artificial, Sistemas Especialistas, também chamados de sistemas baseados em conhecimento, são construídos a partir de bases de informações relacionadas entre si por regras de inferência. A informação e as regras, referentes a áreas restritas e altamente especializadas de conhecimento, são obtida a partir de especialistas que informam como realizam suas tarefas e como chegam a conclusões sobre questões relacionadas à sua especialidade. Áreas em que são usados estes sistemas são, por exemplo, diagnóstico médico e prospecção geológica. O conhecimento que o sistema manipula deve estar adequadamente representado, através de um esquema de representação da informação e das regras. Estes sistemas, para serem utilizados, requerem que o operador seja, por si, versado na área de aplicação, e apto a realizar as perguntas necessárias, de uma maneira que o sistema as reconheça, e, mais importante, que possa interpretar e aplicar a resposta obtida. Conceitos e procedimentos codificados nos sistemas não podem ser indefinidamente explicados, restando sua compreensão como baseada no conhecimento tácito que o operador tem sobre como interpretá-los e agi-los no mundo.

Por este motivo também, além de serem limitados em seu escopo de especialidade, sistemas especialistas lidam com dificuldade com questões baseadas em premissas corriqueiras, disponíveis para pessoas não especialistas, mas que não tenham sido explicitamente codificadas. Este é o conhecimento usualmente chamado de "senso comum". O sistema que abordamos, Cyc, é uma iniciativa que se propõe a coletar todo este senso comum, e torná-lo disponível como uma espécie de infraestrutura de conhecimento básica para outros sistemas especialistas.

Pese embora o interesse prático de um sistema deste tipo, procuramos observar atentamente, por outro lado, como é proposto Cyc e o conhecimento nele inscrito. Seu projeto constitui, claramente, uma proposição epistemológica, construindo um marco materializado do que significa conhecimento e sua manipulação. Para expor mais claramente as posições epistemológicas próprias de Cyc, apresentamos um conjunto de premissas epistemológicas que, trazendo considerações das ciências sociais, provêem um contraste às posições dos programadores de Cyc. As reivindicações de um conhecimento universal e consensual feitas pelos autores de Cyc, então, ganham um contraponto. Em primeiro lugar, o conhecimento que Cyc procura representar e codificar é estritamente declarativo, excluindo assim conhecimentos como aqueles imbuídos no corpo; ademais, deve ser explícito, levando então à necessidade de explicitar conhecimento tácito. Explicitar conhecimento tácito é possível em certa medida, mas há a questão da regressão na medida em que a compreensão e interpretação do mundo apoia-se em uma densa rede de percepções; esta densa rede, tácita e implícita, não é necessariamente explicitada para a vivência cotidiana, sendo expressa apenas quando colocada em questão, na situação própria em que a questão é colocada.

Tácito para o projeto Cyc, no entanto, é a proposição das declarações que constituem sua base de informações como sendo um conhecimento puramente objetivo, eclipsando o sujeito que sabe, e o lugar de onde este conhecimento se dá. Esta elisão do sujeito não coloca em discussão quem enuncia o conhecimento ali contido, um quem composto por um grupo social e cultural muito específico, que é o de engenheiros e pesquisadores acadêmicos, anglófonos, com base territorial nos Estados Unidos, e especialistas de algumas outras áreas. O saber relatado por estas pessoas é válido, mas mostra, sob interrogação cuidadosa, as marcas de sua origem extremamente específica, em franco contraste com as reivindicações de universalidade. Por fim, o projeto ativamente não

discute as implicações políticas de uma base de dados, que alega conter conhecimento comum, no que tange às consequências implicadas para pessoas ou grupos cujo conhecimento não é, em forma ou em conteúdo, compatível com aquele na base, ou que, por outro lado, não tenham acesso ou direito de manipulação desta base.

O projeto Cyc constitui uma ponte epistemológica, identificada pelos participantes do grupo onde foi realizado o trabalho de campo, entre a enunciação de seus próprios saberes, legitimados dentro das práticas científicas e tecnológicas, e o conhecimento humano em uma acepção mais abrangente. A caracterização do humano, dentro do grupo e seguindo uma tradição dentro da Inteligência Artificial, baseia-se no cognitivismo, e essencialmente considera o saber como resultado de um processamento adequado de informação percebida, em uma analogia – não gratuita – com o computador como ferramenta de processamento de dados. Uma outra faceta desta caracterização é a estratégia de projetos de sistemas, utilizada no grupo, baseada em agentes inteligentes. Agentes são conceituados como entes isolados, com um acesso ao mundo dividido em canais de entrada, que enviam sinais informativos sobre o ambiente para o interior do agente, e canais de saída, pelos quais o agente pode agir sobre o mundo. As ações são determinadas internamente, através do processamento dos sinais de entrada e de um estado interno definido. O agente, segundo esta conceituação, está separado do mundo, e seu estado interno inclui, à guisa de conhecimento, uma representação estável, mais ou menos fiel, deste ambiente externo. O humano é subsumido a este modelo, sendo caracterizado explicitamente como um tal agente; sistemas computacionais atribuídos de inteligência são construídos, pelos participantes, a partir deste modelo.

Assim com em relação a Cyc, esta caracterização foi estudada em comparação com caracterizações da agência oriundas das ciências sociais. A agência, absoluta e isolada na acepção corrente no grupo, pode ser vista, alternativamente, como modalizada de várias maneiras dentro do mundo social. A agência propriamente humana é avaliada não de maneira absoluta, mas no questionamento sobre a diferença realizada disponível para apreciação por outras pessoas – uma agência que é julgada dentro dos critérios, e a partir da decisão situada sobre o que conta como preenchendo os critérios, como é o caso da agência histórica, atribuída (com justiça ou não) a certas personagens. Por outro lado, a avaliação social sobre a diferença realizada leva a considerar artefatos como os tecnológicos como possuidores de agência, na medida em que participam das ações e

constituem diferença nestas. Artefatos tecnológicos como os sistemas de IA passam a ser reconhecidos enquanto agentes, mas em um estatuto que não é subsumível ou identificável com a agência humana.

Estas considerações foram colocadas em jogo para compreender a construção de agentes inteligentes no trabalho do grupo. Uma construção de emoção, descrita como característica humana, é imbuída nestes agentes, a partir de uma sistemática própria a partir de uma caracterização oriunda de uma teoria psicológica, uma teoria cognitiva das emoções. Esta construção define emoções como a consequência da avaliação de situações e ações presentes para a pessoa; em linha com a equiparação agente e humano, estas avaliações são transpostas para os agentes artificiais. Um exame mais detido da teoria mostra em primeiro lugar que foi desenvolvida para ser computacionalmente tratável, objetivo que torna mais compreensível a estrutura discreta e catalogável imposta às emoções – quantificadas em 22 – assim como a cobertura completa do campo emocional através de um mapa das situações e regras que produzem a emoção. Os autores da teoria recorrem a procedimentos que já encontramos dentro da IA: emoções são descritas por informação e regras de inferência, e o sujeito que se emociona, e o universo social em que esta emoção ganha significados, é completamente eclipsado. O universo social aparece, sutilmente, como problema, quando os autores abordam temas emocionais não explicáveis pelas construções propostas (como é o caso da sua perplexidade diante do regozijo pela desgraça alheia...)

A ser colocada em uso nos projetos do grupo, esta concepção de emoção e de humano carrega consigo estas marcas da concepção da afetividade como previamente. Os sistemas são pensados, pelos participantes, com o objetivo de propor encontros com as pessoas nos quais o interpretar e o sentir são previamente calculados, ou esta é a expectativa. A teoria das emoções empregada é estendida com regras para determinar – mesmo que seja estatisticamente – a ação adequada em função da emoção calculada. Um dos sistemas apresenta uma animação gráfica em forma de vídeo, em que personagens que são agentes artificiais agem; a ação é construída sobre uma caracterização de personagens em que a intenção é que a avaliação moral sobre cada um seja claramente acessível por quem assistir a animação. Nesta animação, um par binário de personagens identifica uma situação vítima-agressor, e uma avaliação em linha com esta caracterização é

explicitamente esperada; deve-se simpatizar com a vítima reconhecida, deve-se rejeitar o agressor reconhecido.

Uma estruturação semelhante é visível em outra construção realizada no grupo: a da cultura. Novamente, um ensaio sobre uma característica humana é materializado, em sua transposição para a realização computacional, e a mídia de apresentação são animações em forma de vídeo, em que personagens realizam condutas destinadas a serem interpretadas pelo espectador. A cultura é caracterizada como geradora de diferenças, como dando origem a condutas reconhecivelmente diferentes; e como marcador concretizante do abstrato cultural, entra em cena o ritual. Busca-se caracterizar o ritual como uma sequência de ações que segue uma intenção e um plano, ou roteiro, no entanto esta caracterização é sentida como carente de distintividade em relação a outras ações coordenadas e planejadas que povoam o léxico agencial da IA; a solução de compromisso para conferir distintividade é considerar que o ritual é importante por suas ações, não por um objetivo utilitário. Para o sistema, uma animação de ritual é construído como uma cena cotidiana, a de um jantar, encenada duas vezes. Cada encenação tem diferenças na conduta coletiva dos personagens artificiais que decorre de traços considerados prévios para distinguir culturas, tais como atributos individualismo ou coletivismo.

Cultura, como realização própria humana, desperta o interesse para uma realização em artefatos inteligentes – onde ganha o nome de cultura sintética. O fenômeno ritual é eleito, em um efeito sinedóquico, para concretizar esta intenção. A maneira como ritual é formulado dentro deste contexto é interessante; apontado, pela Antropologia, como um momento especial sentido como tal pelas pessoas que o encenam, é aqui, de maneira especular, selecionado pelo construtor do sistema. Uma pitoresca cena de jantar, com um certo cerimonial, é criada em animação; o próprio caráter pitoresco sugere que o narrador é um *outsider*, mas a descrição é complicada pelo fato de que, em última análise, quem responde pelo significado deste ritual é o mesmo construtor. O narrador-programador criou uma imagem prévia, que constitui a seus olhos a distintividade cultural, mas onde está a vivência cultural que preenche aquele momento, algo cerimonial para nossos olhos, com seu significado, além de no próprio construtor? Há um atravessamento curioso, em que o ritual é encenado através de uma interpretação de quem não vive a cosmogonia encenada na performance, e em que o significado íntimo estabelecido pelo desempenhar do ritual não é presente, enquanto que o olhar estrangeiro, ávido da cena visual, está.

Lado a lado com este estabelecimento assimétrico do direito a representar, está uma posição que naturaliza o artefato como tábua rasa, acultural e transcendente, sobre a qual se podem inscrever as contingências culturais.

Este atravessamento é mais explícito em um outro projeto relacionado, que emprega agentes artificiais e animações gráficas, e executado por um consórcio de grupos de pesquisa europeus. Este projeto tem como uma de suas realizações um jogo para computador. O jogo, de caráter pedagógico, transcorre em um cenário de ficção científica em que o jogador interage, em um planeta distante, com seres alienígenas, e com sua cultura peculiar. O objetivo é que crianças europeias possam, através do jogo, passar a conviver melhor com seus colegas refugiados de outros países. Um conjunto complexo de relações de representação, de autoridade e de assimetria são aqui materializados, remetendo a condições históricas em que a tradição de saber europeia discursa sobre o *outro*, oriundo de lugares marcados por sua condição de ex-colônia: quem pode descrever uma cultura, quem são refugiados, quem compreende o outro, quem é pitoresco. Uma aproximação pode ser feita com a investigação, até hoje extremamente atual e suscitadora de debates, realizada por Saïd sobre o orientalismo como construção discursiva, euro-americana, que constitui o objeto Oriente através da demarcação de uma diferença, romantizada e modelizada, que fala sobre um *si* ocidental ao tentar distinguir-se do *outro*. Não por acaso, certamente, o nome do projeto é Orient (acrônimo de *Overcoming Refugee Integration with Novel Technologies*).

Por fim, o último caso parte de um sistema de agentes inteligentes, e gira em torno de um percurso do artefato fora do grupo, em seu encontro com usuários. O sistema consiste de um agente que joga xadrez e comenta sobre o jogo com a pessoa com quem está jogando; o agente pode funcionar, "animar", um robô em forma de gato ou um telefone celular com tela gráfica. A questão colocada era se pessoas conseguiam ver o agente como o "mesmo" nas duas situações, e para investigá-la um experimento que alterava a aparência e outras características de cada implementação do sistema foi proposto. A intenção original da investigação, que era a de determinar atributos que previamente programados levassem ao julgamento de identificação, foi colocada em xeque por resultados contraditórios, em que a presença ou não de características não influenciava estatisticamente a avaliação sim/não. No entanto, os comentários dos usuários, sobre diferenças e semelhanças percebidas, mostra claramente que os usuários observaram

atentamente as características propostas para os agentes, e, mais interessantemente, foram além, discernindo outras características além das projetadas pelos pesquisadores. Examinamos estas respostas para compreender como os usuários julgam traços perceptíveis, sendo provocados a interpretar sinais cujo significado, para eles, não está configurado de maneira estável. Esta interpretação levou os usuários a reconhecer estes traços, contudo a decisão crucial sobre se o traço, como percebido, preenche ou não o critério de identidade depende de um ato cognitivo situado, que não foi compartilhado entre usuários no experimento nem constitui parte de um repertório cultural estabelecido. Identidade, ou "ser o mesmo", é um julgamento não trivial realizado com base em habilidades e procedimentos codificados culturalmente e cuja aplicação é aprendida ao longo da trajetória social da pessoa; ademais, é um julgamento vinculado ao estatuto ontológico do ente do qual se predica a identidade. Entre idas e vindas, o resultado não definido do julgamento não significava déficit ou falha dos pesquisadores, da implementação do sistema nem da compreensão dos usuários: simplesmente foi o resultado do encontro de um artefato rico em características interessantes com um público que vasculhou sua presença disponível e trabalhou intensamente as percepções interpretações ali possíveis.

A presente tese, em suma, abordou a construção de uma série de noções de humano, imbuídas em artefatos tecnológicos realizadas por grupos em um ramo particular da Ciência da Computação que é a Inteligência Artificial, em especial trabalhos relacionados ao grupo em que foi realizado o trabalho de campo. Todos estes artefatos são construídos com a reivindicação de dar conta, ou implementar computacionalmente, aspectos do *ser* humano. A implementação de tais traços, proposta como maneira de tornar estes sistemas mais aptos, seja a realizar tarefas ainda não realizadas de maneira computacional, seja a serem mais amenos à interação com pessoas, também é expressa em termos de equiparar estes sistemas a humanos em termos de contextos de sociabilidade em que pessoas tomam parte – cujo escopo, desta forma, seria ampliado com parceiros artificiais.

A construção destas características é realizada a partir de várias estratégias. Disciplinas cujo objeto de estudo é relevante para a tarefa são mobilizadas, através de seus textos, teorias e também pelo trabalho direto de pesquisadores destas áreas dentro dos projetos em IA. Características humanas são reconfiguradas como capacidades específicas e desta forma objetivadas. Tanto construir artefatos tangíveis como colocá-los em prova

para demonstrar estas capacidades são parte da rotina de trabalho dos pesquisadores do grupo. Novas capacidades são ativamente procuradas como candidatas a aperfeiçoar o estado da arte em artefatos inteligentes.

Os casos apresentados tornam mais visíveis algumas regularidades no processo abordado de eleger características, figurá-las como capacidades e implementá-las em sistemas computacionais. Para compreender estas regularidades, é necessário colocar-se em uma posição de estranhamento em relação aos conceitos e procedimentos da IA, trazendo outras formas de compreender o humano e o social, e desta forma construindo perspectivas alternativas sobre o humano e sobre as figurações escolhidas e produzidas pelos pesquisadores da IA. As características são estabilizadas e depois construídas sistematicamente através de modelos, e estes modelos devem estabelecer claramente um elenco de categorias, relacionar estas categorias através de regras explícitas, e quantificar de alguma maneira estas relações. Dados estes requisitos para a enunciação de características humanas, o relacionamento entre a IA e as disciplinas a que recorre é modulado: apenas proposições teóricas que dêem conta destes requisitos são consideradas adequadas, e é com estas proposições (apenas) que a IA estabelece diálogo. Neste panorama, podemos identificar uma série de circuitos recorrentes de configuração, tais como a teoria que foi proposta, dentro da psicologia, com o objetivo de ser adequada para a computação, e que é empregada pela computação afetiva como fundamento. Talvez o mais claro destes circuitos seja a figuração especular da inteligência, em que a forma de ser própria dos pesquisadores em IA é o modelo para inteligência – "Tudo o que o pesquisador tem de fazer é olhar no espelhe para ver o exemplo de um sistema inteligente" (Russell & Norvig, 1995, p. 3), que conclui subsumindo, dentro do escopo desta formulação ontológica, humanos, robôs e software, como "agentes genéricos".

## *8.1 Agenda para pesquisas futuras*

Este trabalho, realizado com a intenção de propor novas perspectivas e abrir diálogos entre a Inteligência Artificial e outras formas de compreender o humano, pode ser visto como uma etapa em um percurso mais longo de pesquisa. Refinamentos teóricos que

possam dar conta com mais sofisticação dos conceitos manipulados pela IA são importantes, assim como um olhar mais detido sobre o processo de delimitação e legitimação destes conceitos, para que, melhor conhecidas, estas fronteiras possam ser tornadas mais fluidas. Relacionar melhor e mais detalhadamente fundamentos epistemológicos que foram questionados com sua tradição filosófica específica na história do pensamento ocidental também seria relevante, ao situar e propor um vocabulário mais preciso para o intercâmbio que a tese propõe.

Outro ponto que merece um aprofundamento é a genealogia dos objetos e das noções hoje correntes no campo, especificamente na pesquisa realizada no âmbito do Brasil. A disciplina não segue fronteiras nacionais, mas seus recursos fluem de algumas maneiras particulares – e não de outras – dentro dos trabalhos realizados. Que recursos (tais como dicionários semânticos) são tomados para basear projetos e artefatos, e como estes recursos são manipulados e tornados locais, ou não, são um tema de interesse para o estudo desta prática, em especial em uma nação que percorre o caminho do desenvolvimento de sua infraestrutura científica e tecnológica como o Brasil.

Por fim, creio ser importante destacar o que creio ser uma necessidade de permitir multiplicidades dentro do pensamento científico e de desenvolvimento de tecnologia na academia, principalmente no que se relaciona ao entendimento do humano como múltiplo, situado, pleno de processos em andamento. O presente trabalho investe nesta posição, propondo que não é necessário um consenso de perspectivas, mas que o diálogo sim é essencial. Articular de forma mais precisa esta posição, e entretecê-la ao desenvolvimento metodológico, poderia enriquecer e tornar mais interessante esta linha de investigação.

## *8.2 Considerações Finais*

A Informática na Educação, no sentido estrito da produção de sistemas informatizados para o ensino formal, e no sentido mais amplo da construção e mediação, através de tecnologias de comunicação e informação, de instâncias de caráter pedagógico, está implicada estreitamente com os tópicos desenvolvidos nesta tese. Como vimos,

questões como o que significa *conhecimento*, ou como emoções são *apresentadas* e *esperadas* de jovens escolares, ou a posição a partir de que falar do *outro* com estes jovens, são consideradas importantes o suficiente para ganharem a atenção, e serem construídas com o auxílio de uma gama peculiar de formulações que entrelaçam tecnologia computacional e características humanas: a Inteligência Artificial. Mais do que somente cumprir importantes papéis como funcionalidade tecnológica, o faz através da exploração e recriação de traços emprestados do humano, o que lhe confere um interesse especial, e mesmo um encanto próprio.

Consideramos, portanto, que a articulação a ser levada em conta é ainda mais ampla do que a relação entre Informática na Educação, a Inteligência Artificial, e novas estratégias analíticas. Argumentamos que é importante problematizar e situar esta produção em termos sociais e culturais, com a possibilidade de conduzir a uma compreensão mais rica e mais acessível da contribuição que ela tem a oferecer para a sociedade. Não tratamos de invalidar legitimidade desta produção, mas, ao contrário, de gerar novos estatutos de validade, uma validade em perspectiva dentro do universo do humano.

Oferecemos, assim, um ponto de partida para um diálogo entre a tecnologia e a sociedade neste contexto, construído a partir de premissas diferentes que também podem ser levadas em conta. Cremos que a presença destes questionamentos seja importante, não para um consenso monolítico sobre o que é o humano e como devemos pensá-lo na relação com os objetos que criamos, mas para produzir zonas de encontro com estes objetos que sejam ainda mais criativas e, talvez, mais democráticas.

# *Bibliografia*


Abu-Lughod, L.: *Veiled Sentiments: Honor and Poetry in a Bedouin Society*. Updated ed. with a new preface. ed. Berkeley : University of California Press, 1999 — ISBN 0520224736

Adam, A.: *Artificial Knowing : Gender and the Thinking Machine*. Florence, KY, USA : Routledge, 1998 — ISBN 9780415129633 9780203005057

Arnaiz, M. G.: Nutritional Discourse in Food Advertising: Between Persuasion and Cacophony. In: *Anthropology of food, Traditions et identités alimentaires locales*. (Issue 0) (2001)

Ashmore, M. ; Wooffitt, R. ; Harding, S.: Humans and Others, Agents and Things. In: *American Behavioral Scientist* vol. 37 (6) (1994), pp. 733-740

Basalla, G.: *The Evolution of Technology* : Cambridge University Press, 1988 — ISBN 0521296811

Bastide, R.: *O candomblé da Bahia: rito nagô*. São Paulo : Companhia das Letras, 2001 — ISBN 9788535901375

Bell, C. M.: *Ritual: perspectives and dimensions* : Oxford University Press US, 1997 — ISBN 9780195110517

Bloor, D.: The Strong Programme in the Sociology of Knowledge. In: *Knowledge and Social Imagery*. 2. ed. Chicago : University of Chicago Press, 1991. — 1976 — ISBN 0226060969, pp. 3-23

van Breemen, A.: iCat: Experimenting with Animabotics. In: *Proceedings of the Symposium on Robotics, Mechatronics and Animatronics in the Creative and Entertainment Industries and Arts*. Hatfield, UK : The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 2005 — ISBN 1 902956 43 3, pp. 27-32

van Breemen, A. ; Yan, X. ; Meerbeek, B.: iCat: an animated user-interface robot with personality. In: *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. The Netherlands : ACM, 2005 — ISBN 1-59593-093-0, pp. 143-144

Brooks, R. A.: Intelligence Without Reason. In: Steels, L. ; Brooks, R. A. (eds.): *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*. Hillsdale, N.J : Lawrence Erlbaum, 1995 — ISBN 0805815198

Callon, M.: Some elements of a sociology of translation: Domestication of the scallops of St. Brieuc Bay. In: *Power, Action, and Belief: A New Sociology of Knowledge?* London : Routledge & Kegan Paul, 1986 — ISBN 0710208022

Callon, M.: Actor-Network Theory - The Market Test, Published by the Department of Sociology,

Lancaster University, Lancaster LA1 4YL, UK (1997)

Castañeda, C. ; Suchman, L.: Robot Visions, Department of Sociology, Lancaster University (2005)

Collins, H. M.: *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, Mass : MIT Press, 1990 — ISBN 026203168X

Copeland, J. B.: Alan Turing 1912-1954. In: Copeland, J. B. (ed.): *The Essential Turing: Seminal Writings in Computing, Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*. Oxford : Clarendon Press, 2004 — ISBN 0198250797, pp. 5-57

Coulon, A.: *Etnometodologia*. Petrópolis : Vozes, 1995 — ISBN 85-326-1411-6

Crist, E.: Can an Insect Speak?: The Case of the Honeybee Dance Language. In: *Social Studies of Science* vol. 34 (1) (2004), pp. 7-43

Cuba, P.: *Agent Migration between Bodies and Platforms*. Lisboa, Instituto Superior Técnico, MSc Thesis, 2010

Damásio, A. R.: *O erro de Descartes : emoção, razão e o cérebro humano.* São Paulo : Companhia das Letras, 1996 — ISBN 8571645302

Damo, A. S.: *Do Dom à Profissão: uma etnografia do futebol de espetáculo a partir da formação de jogadores no Brasil e na França*. Porto Alegre, Universidade Federal do Rio Grande do Sul, PhD Thesis, 2005

Despret, V.: *Our Emotional Makeup: Ethnopsychology and Selfhood*. New York : Other, 2004 — ISBN 1590510364

Douglas, M.: Deciphering a Meal. In: *Daedalus* vol. 101 (1) (1972), pp. 61-81. — ArticleType: primary_article / Issue Title: Myth, Symbol, and Culture / Full publication date: Winter, 1972 / Copyright © 1972 American Academy of Arts & Sciences

Dourish, P. ; Brewer, J. ; Bell, G.: Information as a cultural category. In: *interactions* vol. 12 (4) (2005), pp. 31-33. — http://www.dourish.com/publications/2005/interactions-information.pdf

Dreyfus, H. L.: *Michel Foucault: Beyond Structuralism and Hermeneutics*. 2. ed. Chicago : University of Chicago Press, 1983 — ISBN 0226163121

Dror, O. E.: Counting the affects: discoursing in numbers. In: *Social Research* vol. 68 (2 (Summer 2001)) (2001), pp. 357-78

eCircus Consortium: *The ORIENT software - ECIRCUS - EU Framework VI Project - Education through Characters with emotional-Intelligence and Role-playing Capabilities that Understand Social interaction*. URL http://www.macs.hw.ac.uk/EcircusWeb/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=50&MMN_position=67:67. - retrieved 2010-12-02

Edwards, D.: Imitation and Artifice in Apes, Humans, and Machines. In: *American Behavioral Scientist* vol. 37 (6) (1994), pp. 754-771

Edwards, P.: From "Impact" to Social Process: Computers in Society and Culture. In: Jasanoff, S. ; Markle, G. E. ; Petersen, J. C. ; Pinch, T. (eds.): *Handbook of Science and Technology Studies*. Rev. ed. ed. Thousand Oaks, Calif : Sage, 2001 — ISBN 0761924981, pp. 257-285

Flick, U.: *Uma Introdução à Pesquisa Qualitativa*. 2. ed. Porto Alegre : Bookman, 2004 — ISBN 9788536304144

Flusser, V.: *Filosofia da Caixa Preta*, 2002

Forsythe, D. E.: Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. In: *Social Studies of Science* vol. 23 (3) (1993), pp. 445-477

Forsythe, D. E.: "It's Just a Matter of Common Sense": Ethnography as Invisible Work. In: *Computer Supported Cooperative Work (CSCW)* vol. 8 (1) (1999), pp. 127-145

Fuller, S.: Making Agency Count: A Brief Foray Into the Foundations of Social Theory. In: *American Behavioral Scientist* vol. 37 (6) (1994), pp. 741-753

Garfinkel, H.: *Studies in Ethnomethodology*. Cambridge, UK : Polity Press, 1984 — ISBN 0745600050

Goodwin, C.: Seeing in Depth. In: *Social Studies of Science* vol. 25 (2) (1995), pp. 237-274

Google TechTalks: *Computers versus Common Sense*, 2006

Grand, S.: *Growing up with Lucy: How to Build an Android in Twenty Easy Steps*. London : Phoenix, 2004 — ISBN 0753818051

Grint, K. ; Woolgar, S.: Computers, Guns, and Roses: What's Social about Being Shot? In: *Science, Technology & Human Values* vol. 17 (3) (1992), pp. 366 -380

Guha, R. V. ; Lenat, Douglas: CYC: A Midterm Report. In: *AI Magazine* (1990)

Guimarães Jr., M. J. L.: *The configuration of avatars: How users, researchers, and designers write a technology*, Brunel University, Londres, Tese de doutorado, 2005

Harmon, P.: *Expert Systems: Artificial Intelligence in Business*. New York : Wiley, 1985 — ISBN 0471808245

Hayles, K. N.: Computing the Human. In: *Theory Culture Society* vol. 22 (1) (2005), pp. 131-151

Hess, D.: Ethnography and the Development of Science and Technology Studies. In: Atkinson, P. A. ; Delamont, S. ; Coffey, A. J. ; Lofland, J. ; Lofland, L. H. (eds.): *Handbook of Ethnography* : Sage Publications Ltd, 2007 — ISBN 1412946069, pp. 234-245

Hofstede, G.: Motivation, Leadership and Organization: Do American Theories Apply Abroad. In: *Organizational Dynamics* vol. 9 (1) (1980), pp. 42-63

Interrante, J.: The Road to Autopia: the automobile and the spatial transformation of the american culture. In: Lewis, D. L. ; Goldstein, L. (eds.): *The Automobile and American Culture* : University of Michigan Press, 1983, pp. 89-104

Jennings, N. R. ; Sycara, K. ; Wooldridge, M.: A Roadmap of Agent Research and Development. In: *Autonomous Agents and Multi-Agent Systems* vol. 1 (1) (1998), pp. 7-38

Karmiloff-Smith, A. ; Inhelder, B.: If you want to get ahead, get a theory. In: *Cognition* vol. 3 (3) (1975), pp. 195-212

Knorr Cetina, K.: Laboratory Studies: The cultural approach to the study of science. In: Jasanoff, S. ;

Markle, G. E. ; Petersen, J. C. ; Pinch, T. (eds.): *Handbook of Science and Technology Studies*. Rev. ed. ed. Thousand Oaks, Calif : Sage, 2001 — ISBN 0761924981, pp. 140-166

Latour, B.: *Jamais fomos modernos : ensaio de antropologia simétrica*. São Paulo : Editora 34, 1994 — ISBN 8585490381

Latour, B.: *Ciência em Ação: como seguir cientistas e engenheiros sociedade afora*. São Paulo : UNESP, 2000

Latour, B.: *A Esperança de Pandora*. Bauru : EDUSC, 2001 — ISBN 8574600628

Lee, N. ; Brown, S.: Otherness and the Actor Network. In: *American Behavioral Scientist* vol. 37 (6) (1994), pp. 772 -790

Lévy, P.: *As Tecnologias da Inteligência* : 34, 1993

Li, D.: *Artificial Intelligence with Uncertainty*. Boca Raton : Chapman & Hall/CRC, 2008 — ISBN 9781584889984

Lofland, J. ; Snow, D. A. ; Anderson, L. ; Lofland, L. H.: *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. 4. ed. : Wadsworth Publishing, 2005 — ISBN 0534528619

Luger, G. F.: *Artificial intelligence: structures and strategies for complex problem solving* : Addison-Wesley, 2002 — ISBN 9780201648669

Marcus, J.: Orientalism. In: Atkinson, P. A. ; Delamont, S. ; Coffey, A. J. ; Lofland, J. ; Lofland, L. H. (eds.): *Handbook of Ethnography* : Sage Publications Ltd, 2007 — ISBN 1412946069, pp. 109-117

Maturana, H. ; Varela, F.: *A Árvore do Conhecimento: as bases biológicas da compreensão humana*. São Paulo : Palas Athena, 2001

Mccarthy, J. ; Hayes, P. J.: Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: Meltzer, B. ; Michie, D. (eds.): *Machine Intelligence 4* : Edinburgh University, 1969, p. 463--502

McCulloch, W. ; Pitts, W.: A logical calculus of the ideas immanent in nervous activity. In: *Bulletin of Mathematical Biology* vol. 5 (4) (1943), pp. 115-133

Merton, R.: The Normative Structure of Science. In: *The sociology of science : theoretical and empirical investigations*. Chicago : University of Chicago Press, 1973 — ISBN 9780226520919, pp. 267-278

Naphade, M. ; Smith, J. R. ; Tesic, J. ; Chang, S.-F. ; Hsu, W. ; Kennedy, L. ; Hauptmann, A. ; Curtis, J.: Large-Scale Concept Ontology for Multimedia. In: *IEEE MultiMedia* vol. 13 (3) (2006), pp. 86-91

Ortony, A. ; Clore, G. L. ; Collins, A.: *The cognitive structure of emotions* : Cambridge University Press, 1990 — ISBN 9780521386647

Panton, K. ; Matuszek, C. ; Lenat, D. ; Schneider, D. ; Witbrock, M. ; Siegel, N. ; Shepard, B.: Common Sense Reasoning-From Cyc to Intelligent Assistant. In: *Lecture notes in computer science* vol. 3864 (2006), p. 1

Pearl, J.: Probabilistic and Qualitative Abduction. In: *AAAI Spring Symposium on Abduction*. Stanford : AAAI Press, 1991, pp. 155-158

Pearl, J.: Decision making under uncertainty. In: *ACM Comput. Surv.* vol. 28 (1) (1996), pp. 89-92

Peirano, M.: *A Análise Antropológica de Rituais*. Brasília : Departamento de Antropologia - UNB, 2000

Peirano, M.: *Temas ou Teorias? O estatuto das noções de ritual e de performance*. Brasília : Departamento de Antropologia - UNB, 2006

Pfaffenberger, B.: Fetishised Objects and Humanised Nature: Towards an Anthropology of Technology. In: *Man, New Series*. vol. 23 (2) (1988), pp. 236-252. — http://www.jstor.org/stable/2802804

Pfaffenberger, B.: Social Anthropology of Technology. In: *Annual Review of Anthropology* vol. 21 (1992), pp. 491-516. — http://www.jstor.org/stable/2155997

Picard, R. W.: *Affective Computing*. Cambridge, Mass : MIT Press, 1997 — ISBN 0262161702

Porayska-Pomsta, K. ; Pain, H.: Exploring Methodologies for Building Socially and Emotionally Intelligent Learning Environments. In: *Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments (SEILE), ITS*. Maceió, Brasil, 2004

Riskin, J.: The Defecating Duck, or, the Ambiguous Origins of Artificial Life. In: *Critical Inquiry* vol. 29 (4) (2003), pp. 599-633. — http://www.stanford.edu/dept/HPS/DefecatingDuck.pdf

Rumelhart, D. E. ; Hinton, G. E. ; Williams, R. J.: Learning representations by back-propagating errors. In: *Nature* vol. 323 (6088) (1986), pp. 533-536

Russell, S. J. ; Norvig, P.: *Artificial Intelligence: A Modern Approach*, *Prentice Hall series in artificial intelligence*. Englewood Cliffs, N.J : Prentice Hall, 1995 — ISBN 0131038052

Russell, S. J. ; Norvig, P.: *Artificial Intelligence: A Modern Approach*. 3. ed. : Prentice Hall, 2010 — ISBN 9780136042594

Saïd, E. W.: *Orientalismo: o Oriente como invenção do Ocidente*. São Paulo : Companhia das Letras, 1990 — ISBN 85-7164-133-1

Sá, J. G. da S.: *No mesmo galho: ciência, natureza e cultura nas relações entre primatólogos e primatas*, Museu Nacional, UFRJ, 2006

Shapin, S.: *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton : Princeton University Press, 1985 — ISBN 0691024324

Sloman, A.: Damasio, Descartes, Alarms and Meta-management. In: *In Proceedings International Conference on Systems, Man, and Cybernetics (SMC98* : IEEE, 1998, pp. 2652–7

Sontag, S.: *Diante da Dor dos Outros*. São Paulo : Companhia das Letras, 2003 — ISBN 8535903984

Star, S. L. ; Strauss, A.: Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. In: *Computer Supported Cooperative Work (CSCW)* vol. 8 (1) (1999), pp. 9-30

Steels, L.: Fifty Years of AI: From Symbols to Embodiment - and Back. In: Lungarella, M. ; Iida, F. ; Bongard, J. ; Pfeifer, R. (eds.): *50 years of artificial intelligence : essays dedicated to the 50th anniversary of artificial intelligence*. Berlin; New York : Springer, 2007 — ISBN 9783540772958, pp. 18-28

Steels, L. ; Brooks, R. A. (eds.): *The Artificial Life Route to Artificial Intelligence: Building Embodied,*

*Situated Agents*. Hillsdale, N.J : Lawrence Erlbaum, 1995 — ISBN 0805815198

Stewart, A.: *The Ethnographer's Method*. Thousand Oaks, Calif : Sage Publications, 1998 — ISBN 0761903933

Suchman, L.: Embodied Practices of Engineering Work. In: *Mind, Culture, and Activity* vol. 7 (1) (2000), pp. 4-18

Suchman, L.: *Human-Machine Reconfigurations: Plans and Situated Actions*. 2. ed. Cambridge : Cambridge University Press, 2007 — ISBN 9780521858915

Tambiah, S. J.: A Performative Approach to Ritual. In: *Culture, Thought, and Social Action: An Anthropological Perspective*. Cambridge, Mass : Harvard University Press, 1985 — ISBN 0674179692, pp. 123-135

Taylor, M. E. ; Matuszek, Cynthia ; Klimt, B. ; Witbrock, M.: Autonomous classification of knowledge into an ontology. In: *Proceedings of the Twentieth International FLAIRS Conference (FLAIRS 2007)* : AAAI Press, 2007 — ISBN 978-1-57735-319-5, pp. 140-145

Teubner, G.: Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law. In: *Journal of Law and Society* vol. 33 (4) (2006), pp. 497-521

Thompson, C.: *Making Parents: The Ontological Choreography of Reproductive Technologies*, *Inside technology*. Cambridge, Mass : MIT, 2005 — ISBN 0262201569

Turing, A. M.: Computing Machinery and Intelligence. In: *Mind* vol. LIX (236) (1950), pp. 433 -460

UNDP: *United Nations Development Program - Human Development Report 2001*. New York : Oxford University Press, 2001. — http://hdr.undp.org/en/media/completenew1.pdf — ISBN 0195218361

Vicente, A. de ; Pain, H.: Informing the Detection of the Students' Motivational State: An Empirical Study. In: *Proceedings of the 6th International Conference on Intelligent Tutoring Systems* : Springer-Verlag, 2002. — DOI 10.1007/3-540-47987-2_93 — ISBN 3-540-43750-9, pp. 933-943

Weizenbaum, J.: ELIZA - a computer program for the study of natural language communication between man and machine. In: *Commun. ACM* vol. 9 (1) (1966), pp. 36-45

Wiener, N.: *Cibernética; ou Controle e comunicação no animal e na máquina*. São Paulo : Polígono e Universidade de São Paulo, 1970

Wild, R. ; Cuba, P. ; Prada, R. ; Biazus, M. C.: Julgando por aparências, buscando diferenças: o jogo da interpretação entre humanos e agentes artificiais. In: *Anais do XX Simpósio Brasileiro de Informática na Educação*. João Pessoa, 2010 — ISBN 2176-4301

Wild, R. ; Maurente, V. ; Maraschin, C. ; Biazus, M. C.: "Coisas que pessoas sabem": Computação e territórios do senso comum. In: *Scientia Studia* vol. 9 (1) (2011)

Witbrock, M. ; Matuszek, Cynthia ; Brusseau, A. ; Kahlert, R. C. ; Fraser, C. B. ; Lenat, D.: Knowledge begets knowledge: Steps towards assisted knowledge acquisition in Cyc. In: *Proc. of the AAAI 2005 Spring Symposium on Knowledge Collection from Volunteer Contributors* : AAAI Press, 2005, pp. 99-105