Ciências & Cognição 2014; Vol 19(3) 416-426 <http://www.cienciasecognicao.org> © Ciências & Cognição
Submetido em 17/05/2013 | Revisado em 18/06/2014 | Aceito em 21/06/2014 | ISSN 1806-5821 – Publicado on line em 01/12/2014

Mentes, Metas e Máquinas: um desafio para a consciência artificial

Minds, Goals and Machines: a challenge for artificial consciousness

Samuel de Castro Bellini-Leite

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

Resumo

A Teoria do Espaço de Trabalho Global pretende descrever funcionalmente as tarefas da consciência. Atualmente existem tentativas de aplicação desta teoria cognitiva da consciência em máquinas, entretanto, esta teoria depende, essencialmente, da noção de metas para determinação de hierarquia do conteúdo na consciência. Este trabalho analisa características de metas e a diferença entre *metas genuínas* e *metas parciais*. Propõe-se que para que hajam *metas genuínas* em um sistema é preciso que este seja líquido (capaz de se adaptar a mudanças) sendo composto de microagentes com interesse em conjunto de sobrevivência e replicação. Sugere-se que duas correntes da Inteligência Artificial atual podem vir a criar agentes com metas genuínas: a nanorrobótica e a vertente das máquinas híbridas. Por essa análise, acredita-se que uma aplicação fidedigna da Teoria do Espaço de Trabalho Global depende primeiramente da existência de sistemas artificiais com metas genuínas.

Palavras-chave: Teoria do Espaço de Trabalho Global; metas; consciência em máquinas.

Abstract

The Global Workspace Theory attempts to describe conscious tasks functionally. Also, there are attempts of applying such cognitive theory to machine consciousness. However, the theory depends essentially on the notion of goals for determining the hierarchy of content in consciousness. This work analyzes characteristics of goals, differentiating “genuine goals” from “partial goals”. This work proposes that in order for a system to have genuine goals it must be liquid, capable of adapting to changes, being composed of micro-agents with the common interest of survival and replication. We propose that two branches of Artificial Intelligence could be capable of creating agents with genuine goals in the future: nanorobotics and hybrid machinery. Because of such analysis we believe a trustworthy application of Global Workspace Theory depends first on the existence of artificial systems with genuine goals.

Keyword: Global Workspace Theory; goals, machine consciousness

Autores de Correspondência:

S.C. Bellini-Leite - Av. Augusto de Lima, n.134, apto. 806, Centro, Belo Horizonte, MG, Brasil. E-mail para correspondência: samuelcblpsi@gmail.com

1. Introdução

A Teoria do Espaço de Trabalho Global é uma das teorias mais influentes sobre o funcionamento da consciência, e segue a vertente da Ciência Cognitiva e do Funcionalismo. Como é comum às teorias cognitivas, existem propostas de implementação desta em máquinas, com o objetivo de, no longo prazo, criar máquinas que possam ter no mínimo alguns aspectos da consciência humana (Baars & Franklin, 2007; Franklin & Patterson Jr, 2006; Madl, Baars & Franklin, 2011; Silva & Gudwin, 2011). Este trabalho não objetiva comentar como esta teoria pretende ser implementada, mas sim discutir alguns aspectos filosóficos da Teoria do Espaço de Trabalho Global para comentar sobre um problema que pode ser encontrado na aplicação desta em máquinas: a implementação de metas.

Seria possível implementar um sistema do Espaço de Trabalho Global em uma máquina? Para responder a esta pergunta, primeiramente é preciso lembrar que existem os problemas dos *Qualia*, das emoções, da subjetividade, que são problemas clássicos enfrentados por qualquer teoria cognitiva, mas que por já serem conhecidos não serão tratados nesse texto. A Teoria do Espaço de Trabalho Global depende essencialmente da noção de meta. Assim, discutir-se-á algumas dificuldades filosóficas para implementação de metas em máquinas. Para tanto, o trabalho está dividido em quatro seções, além desta breve introdução. Na segunda seção, faz-se uma breve revisão de alguns conceitos centrais da Teoria do Espaço de Trabalho Global. Na seção seguinte, serão brevemente apresentadas algumas ideias sobre a origem das metas em seres vivos e algumas propriedades de sistemas líquidos. Na quarta seção, por fim, uma análise sobre diferenças entre "metas genuínas" e "metas parciais" será esboçada, a fim de engendrar uma conclusão.

2. A Teoria do Espaço de Trabalho Global

A Teoria do Espaço de Trabalho Global foi desenvolvida por Bernard Baars, sendo proposta no livro *A Cognitive Theory of Consciousness* (1988). Psicólogo cognitivista, Baars ilustra o Espaço de Trabalho Global com uma analogia de um comitê de especialistas formando uma assembleia na qual nenhum elemento em si consegue resolver o problema desejado sozinho. Como cada especialista tem sua própria forma de compreender e de comunicar o problema, há de início um problema de comunicação entre eles. A solução emerge da utilização de um quadro negro para publicar mensagens globais, para que cada especialista possa entender o que os outros estão fazendo. Como Baars (1988, p. 87, tradução nossa) explica:

> Um passo importante para resolver esse problema de comunicação é pela publicação de uma mensagem global em um grande quadro negro na frente do auditório, para que em princípio qualquer um possa ler a mensagem e reagir. De fato, a mensagem seria lida apenas por especialistas capacitados para entendê-la ou entender partes dela, mas ninguém pode saber, antes do tempo chegar, quem serão esses especialistas, de forma a se tornar necessário que a mensagem seja potencialmente disponível para qualquer um na plateia. A qualquer momento, um número de especialistas pode tentar enviar mensagens globais, mas o quadro negro não pode acomodar todas as mensagens de uma mesma vez – mensagens diferentes serão comumente mutualmente contraditórias. Dessa forma, alguns dos especialistas poderão competir por acesso ao quadro negro, e alguns poderão cooperar em um esforço para publicar a mensagem global[1].

Baars (1988) acrescenta que esse tipo de situação possui vantagens e desvantagens. Um quadro negro não será usado quando o problema é simples e conhecido ou quando uma ação rápida é necessária. Mas é uma forma efetiva de resolver problemas quando há necessidade de cooperação entre fontes distintas de conhecimento, com tempo e vantagens em decidir sobre as possibilidades.

Esta metáfora de Baars (1988) pode incorporar várias características da consciência. Apenas especialistas com certo grau de ativação, nesse caso, com informação relevante para ser publi-

cada terão chances de acesso ao quadro negro. A metáfora pode incorporar a hipótese da novidade, acrescentando a regra da prioridade para conteúdos inovadores para serem publicados, já que conteúdos antigos não seriam vantajosos para publicação. Ela é coerente com o fato da consciência estar relacionada a uma parcela do conteúdo informacional e não a toda informação disponível no cérebro e no mundo.

A linguagem metafórica proposta por Baars (1988) é puramente ilustrativa, sendo uma forma de esclarecer e ajudar o leitor a imaginar a forma de processar desses sistemas; o uso puro de um vocabulário científico pode dificultar a construção dessa imagem clara.

Trocando os membros da assembleia na metáfora por processadores inconscientes especializados e o quadro negro por uma memória de trabalho denominada Espaço de Trabalho Global, Baars (1988) chega a um modelo simples, capaz de abarcar contrastes realizados entre processadores inconscientes e conscientes. Um Espaço de Trabalho Global realiza uma troca de informação que permite a interação de sistemas inconscientes especializados do sistema nervoso. Assim, é um sistema de difusão ("*broadcasting*") da informação para o cérebro. Diversos especialistas inconscientes podem competir ou cooperar para obter acesso ao Espaço de Trabalho Global. Após obtenção de tal acesso, podem transmitir suas informações para todos os outros sistemas.

A teoria abarca a afirmação de processos conscientes serem computacionalmente menos eficientes do que processos inconscientes, pois computações processadas por um Espaço de Trabalho Global exigem ativação de todos os processadores relevantes. Como cada ação precisa de um relativo consenso entre os processadores relevantes, muito mais tempo é gasto em comparação a processos especializados e inconscientes em uma função específica, cujo algoritmo já é conhecido.

Como a mensagem é globalizada, ela precisa atingir todos, ou a maioria dos sistemas inconscientes, utilizando algum tipo de código entendido por todos. Isso incorpora a característica da multimodalidade dos processos conscientes em contraste com a especificidade dos processadores inconscientes. Essa globalização também permite maior relacionamento entre conteúdos, algo difícil para processadores inconscientes com dificuldade de comunicação.

A consistência interna da consciência também é incorporada à teoria. A mensagem presente no Espaço de Trabalho Global em certo momento precisa ser internamente consistente, pois é necessário certo consenso entre os especialistas, ou uma força maior de um grupo que esteja competindo com outro para o acesso. Portanto, existe uma competição ou cooperação para acesso ao espaço de trabalho que precisa globalizar uma mensagem coerente. Esta consistência é evidenciada pelo fato de o olhar humano, por exemplo, enxergar apenas uma forma de uma ilusão de ótica por vez, ou a forma real ou a forma ilusória. Da necessidade de mensagens únicas e coerentes emerge naturalmente a relativa serialidade dos processos conscientes, pois cada mensagem é globalizada uma por vez.

Como a mensagem precisa ser internamente coerente, ela é limitada a poucos itens, pois os itens incompatíveis com a mensagem não podem ser globalizados. O uso do Espaço de Trabalho Global é, portanto, reservado para solucionar apenas os problemas que não podem ser resolvidos por processadores inconscientes especializados funcionando isoladamente.

Assim o Espaço de Trabalho Global possui três características centrais. Ele promove disponibilidade global, de mensagens internamente consistentes, que precisam ser informativas para o sistema.

Um conceito central para a Teoria do Espaço de Trabalho Global é o *contexto*, um grupo de processadores inconscientes com acesso privilegiado à consciência, um sistema que molda a experiência consciente sendo ele próprio inconsciente. A palavra *contexto* normalmente é utilizada para se referir ao ambiente, o qual afeta a cognição. Apesar disso, Baars (1988) utiliza a palavra para se referir às estruturas internas inconscientes, pois o contexto do mundo não influencia nossa cognição se não estiver representado de alguma forma internamente; assim, o psicólogo reduz o contexto ambiental a estruturas inconscientes chamadas também de contexto. Essas estruturas cognitivas estão relacionadas a outras tradicionais da ciência cognitiva, como redes semânticas, planos, roteiros e *frames*. Porém, um novo conceito é necessário para se referir às representações inconscientes agindo para influenciar uma experiência consciente[2].

Um exemplo simples de um *contexto* é um sistema inconsciente realizando previsões em tempo real para compensar a relação em constante mudança com a gravidade e com a paisagem visual. Ao entrar em um barco, de início, o horizonte parece balançar, mas rapidamente percebe-se que o barco está se movimentando. Após algum tempo na água, ao voltar para a terra-firme, percebe-se o mundo como se este estivesse balançando novamente, até que o sistema comece a acertar as previsões. Assim, o mundo é percebido de forma estável apenas quando as previsões são bem sucedidas. Essas previsões são completamente inconscientes, e enquanto estiverem sendo bem sucedidas elas não mostram sua existência.

O fenômeno do *Priming* evidencia esses contextos. O *Priming* acontece quando uma resposta a um estímulo é influenciada pelos estímulos anteriores. Esses efeitos ocorrem em todas as modalidades sensoriais. No caso da linguagem, um bom exemplo é de uma pessoa lidando com frutas. Esta, ao ler a palavra 'manga', automaticamente lembra-se da fruta manga. Mas se esta pessoa estiver lidando com camisas e ler a palavra 'manga', ela, do mesmo modo, imaginará a parte de uma camisa que cobre os braços. Isso evidência contextos representando conteúdos recentes inconscientemente. Ao ouvir uma música ou uma conversa, o indivíduo fica consciente apenas de cada parte de uma vez, mas se cada parte não estivesse sendo relacionada a um todo inconsciente, o mesmo não conseguiria compreender mudanças sintáticas relevantes na fala ou distinguir mudanças interessantes em uma música, como a entrada de um refrão (Baars, 1988).

Baars (1988) resume os grupos de contextos em quatro: contextos de percepção e de imagens mentais; contextos conceituais; contextos de meta; e contextos de comunicação compartilhados entre falantes da mesma conversa.

Os contextos mencionados até agora estão, de modo mais direto, relacionados às expectativas. O contexto que se relaciona diretamente com as intenções é chamado por Baars (1988) de *contexto de meta*. Assim como o conhecido psicólogo Maslow (1970), Baars (1988) trabalha com uma hierarquia de metas. Em um dado momento, as metas são ordenadas de acordo com sua relevância, sendo que as mais relevantes predominam sobre as outras. Por exemplo, num dado momento, a meta de um indivíduo de ler um texto pode estar no topo da hierarquia, mas se sua casa pegar fogo, certamente outra meta irá dominar. O fluxo da consciência pode ser considerado um fluxo de experiências criadas pela dinâmica de diversos contextos de meta, cada um tentando tornar consciente, através da competição pelo acesso ao Espaço de Trabalho Global, aquilo que garantirá progresso em relação à sua meta.

A noção do que é mais informativo será consciente, dependente, portanto, da ideia de um contexto de meta e da hierarquia desses contextos. De toda informação disponível no mundo e na mente humana, apenas aquelas mais próximas às metas se tornarão conscientes. Em máquinas, as metas são impostas por um programador, ou mesmo quando não são completamente impostas, são descritas como regras de prioridade de ações que uma máquina deve exercer para garantir seu objetivo. A fim de elucidar que as metas em organismos vivos carregam uma diferença, tratar-se-á, na seção seguinte, da origem das metas dos organismos.

3. Sistemas Líquidos compostos de microagentes

De acordo com Dennett (1991), antes do surgimento da vida não havia razão, apenas causas, nada possuía propósito. Isto ocorre até o surgimento dos replicadores simples, os quais também não possuíam interesse propriamente dito, como humanos, mas pode-se atribuir-lhes certo tipo de interesse (seguindo aspectos de uma perspectiva de evolução centrada nos genes de Dawkins [1976]). Para que estes replicadores exerçam sua função, seu ambiente precisa possuir condições apropriadas. Quando o comportamento de uma entidade visa anular sua própria decomposição, ela começa agir em termos do que é "bom, ruim e neutro para ela". Dessa forma, ela cria "interesses". Quando a entidade começa a ter interesses, o mundo começa a criar razões para esses interesses, mesmo que a entidade não reconheça. Por exemplo, a razão de um agente buscar o calor é porque isto será bom para o funcionamento de seu sistema.

De acordo com Dennett (1991), as razões vieram antes das criaturas que as reconhecem.

Um dos problemas para os primeiros seres que enfrentaram problemas era reconhecer e agir sobre as razões que o mundo criou a partir dos interesses deles. Para Dennett (1995), desde o início da agência o preço para fazer algo é o de correr o risco de errar. Os primeiros erros foram de cópia. O filósofo considera isto um erro porque existe o preço de errar, diminuição da capacidade de replicar ou até mesmo o fim da linha de reprodução. Antes deste momento não existia a oportunidade de errar, não faria sentido julgar que um processo foi errado; todo erro é subsequente a este primeiro processo. Com o tempo o processo de cópia foi se aperfeiçoando, mas ainda é essencial para a evolução que tal processo não seja rígido; isto porque como Dawkins (1986) deixou claro, as pequenas divergências que podem ocorrer do padrão são exatamente o que gera a mudança e a novidade na vida. De acordo com Dennett (1995), os erros e acertos são relativos à adaptação do sistema a certo meio, e apenas após o sucesso de tais sistemas é possível dizer que um acertou e outro errou, porém até mesmo o que se afirma ter errado pode ser um sistema mais adaptável em potencial, ou seja, algum dia pode vir a ser mais apto do que o anterior.

Quando se tem o objetivo de autorreplicação, criar fronteiras é importante. O organismo não pode ter o objetivo de preservar todo o mundo, ele precisa preservar apenas a si mesmo, tornando-se egoísta. O egoísmo, portanto, é uma das marcas da vida, e também do início da agência. Nesta tarefa de criar barreiras é preciso começar a distinguir os invasores dos amigos, problema que foi resolvido pela criação de formas e detectores de formas. Esta tarefa é essencial para ingestão, excreção, respiração, transpiração e para o sistema imune, ou seja, é essencial para o organismo (Dennett, 1991).

Sumarizando, para Dennett (1991), a mente humana e todas as propriedades que decorrem desta, como a noção de intencionalidade, assim como a noção de meta, aqui investigada, surge da interação de inúmeros microagentes que são herdeiros dos primeiros replicadores.

Uma teoria que também pode ajudar a entender a origem biológica das metas é a teoria do papel da auto-organização na Evolução. Esta teoria pode explicar um pouco mais sobre como a ordem surge. Stuart Kauffman, um biólogo que trabalha com sistêmica, foi um dos pioneiros desta ideia. De início, as tendências da auto-organização na evolução pareciam ser contraditórias às ideias da seleção natural, mas a posição seguida neste trabalho será – de acordo com os raciocínios de Pereira Jr., Paleri, Costa & Guimarães (2004), Kauffmann (1991, 1993) e Dennett (1995) – a de que o fenômeno da auto-organização em sistemas biológicos é completamente compatível e até mesmo complementar à teoria da seleção natural darwinista.

Para entender as novidades propostas por Kauffman e como elas podem ser compatíveis com a seleção natural é preciso esclarecer brevemente o que é auto-organização. Pode-se delimitar um sistema a partir de qualquer agrupamento de elementos exercendo relações entre si em um dado espaço e tempo. Todo sistema que não for fechado é organizado tanto por hetero quanto por auto-organização. Contudo, pode existir em um sistema maior influência de um dos dois tipos de organização. Na hetero-organização há predominância de fatores exógenos atuando no sistema, já na auto-organização o que predomina são fatores endógenos (Pereira Jr. & Pereira, 2010).

Independentemente de qual sistema a realiza, a auto-organização parece ter certas características. Pereira Jr. & Pereira (2010) citam quatro delas: a espontaneidade – a relação entre os elementos traz novidades não contidas no início do processo; possuem a característica de derivar seus padrões de organização das relações internas entre seus componentes e apresentar respostas construtivas às perturbações externas; apresentam distintos níveis de organização os quais estabelecem relações de *feedback* entre si formando uma causalidade circular; e, por fim, apresentam aspectos de não linearidade, desproporção entre magnitude de causas e efeitos no sistema, gerando um "efeito borboleta", ou seja, pequenas mudanças, mesmo que locais, repercutem por todo o sistema, mesmo que ele volte a certo equilíbrio, por um momento, esta pequena mudança é refletida de forma global.

Capturando a essência do conceito, Debrun (1996, p. 4) argumenta que "uma organização ou forma é auto-organizada quando produz a si própria". Para o autor, apesar da forma auto-organizada depender de seus elementos constituintes, estes não determinam mecanicamente o processo a ser desenrolado por base deles. O sistema emergente na auto-organização tem origem no

próprio processo e não nas condições de partida. Assim, esta herda um começo, mas este apenas fornece uma orientação ou impulso numa certa direção. Ainda de acordo com Debrun (1996), o principal motor da auto-organização reside na interação entre elementos realmente distintos e "soltos"; quanto maior a discrepância entre a forma final da interação e a soma das influências, maior é o nível de auto-organização.

O estudo da auto-organização requer algumas novidades de pensamento e método. Para entender um fenômeno complexo, a ciência tradicional divide este em partes cada vez menores para que estas sejam mais bem compreendidas, com o objetivo final de somar as partes e ter uma imagem do todo. O uso quase pejorativo do termo reducionismo é no mínimo equivocado; afinal, do início do século XX até o começo do XXI houve saltos imensos na quantidade de conhecimento e compreensão do mundo baseado nesse método. De toda forma, no final do século XX foram descobertos fenômenos que acontecem no sistema quando analisado como um todo, os quais não são previsíveis estudando apenas os seus elementos. Um dos motivos para a defesa do reducionismo é a dificuldade metodológica para o estudo de sistemas. Contudo, o recurso tecnológico atual permitiu um avanço metodológico inovador para o estudo de sistemas, como por exemplo, o uso de modelagem computacional e modelos matemáticos. A partir destes recentes métodos é possível a análise do comportamento de sistemas como um todo, sem a necessidade de isolar elementos. Como ambos os tipos de pensamentos possuem métodos úteis e não são conflitantes, devem ser complementares.

Para chegar a sua compreensão de seleção natural e auto-organização, Kauffman utiliza uma visão sistêmica e métodos computacionais. Kauffman (1991) afirma que modelos matemáticos ajudam na compreensão de sistemas complexos de processamento em paralelo. Para estudar o comportamento de milhares de elementos interconectados, o biólogo usa uma classe de sistemas chamados de redes booleanas aleatórias. Nestas redes, cada variável é binária e regulada por outras que servem como *inputs*. O comportamento dinâmico de cada variável é controlado por uma função Booleana. Uma função Booleana 'OR', faz a variável responder de forma ativa se alguma dentre suas variáveis de *inputs* está ativa. Uma função 'AND' faz a variável ficar ativa apenas se todos os seus *inputs* estiverem ativos.

Kauffman (1991) discute versões matematicamente idealizadas de sistemas biológicos chamadas de redes booleanas NK autônomas e aleatórias. A rede consiste de elementos (N) ligados por um número (K) de *inputs*. As funções booleanas e o valor das variáveis começam aleatoriamente. São autônomas porque os *inputs* estão dentro do sistema. Porém ao deixar estas redes sem perturbações externas elas permanecem em um de seus ciclos de estado, constantemente. Se a rede for perturbada, sua trajetória pode mudar. Perturbações mínimas são mudanças no valor das variáveis, e perturbações estruturais são mudanças no número de *inputs* (K) por elemento (N) ou mudanças nas funções booleanas.

Em redes com N = K, ou seja, todos os elementos estão interligados, o comportamento é simples e caótico, a sucessão de um estado para outro é completamente aleatória. Estas redes exibem sensibilidade inicial máxima, pois qualquer perturbação pode mudar sua trajetória. Outro sinal de desordem é que quanto mais elementos são incluídos, o tamanho dos ciclos de estado aumenta de forma exponencial, não há padrão ou ordem (Kauffman, 1991).

O autor explica, ainda, que redes com K > 3, ou seja, três ou mais *inputs* por elemento, mantêm comportamentos caóticos como das redes N = K. Contudo, quando se chega a redes K = 2 as redes começam a exibir ordem coletiva espontânea. Os ciclos de estado deste tipo de rede se mantêm estável com quase todas as perturbações mínimas e as perturbações estruturais modificam seu comportamento apenas ligeiramente. A ordem surge sem uma organização precisar ser forçada no sistema, é espontânea. As redes possuem qualidade homeostática, normalmente retornam para seus ciclos de estado padronizados depois de perturbações. Estas redes apresentam características semelhantes as da auto-organização e a homeostase é uma propriedade de todas as coisas vivas.

Kauffman (1991), seguindo a sugestão de Cristopher Langton, classifica o comportamento dessas redes em três tipos, sólido, gasoso e líquido. Uma rede sólida é extremamente ordenada, mas por ser tão rígida, qualquer mudança gera problemas graves para o sistema (como um programa computacional tradicional), ela é inca-

paz de resistir a perturbações, e não se adapta. O sistema gasoso é caótico (como as redes N = K descritas) simples e com mudanças de estado aleatórias. Já as redes líquidas são ordenadas, porém podem chegar à beira do caos. Sistemas líquidos apresentam características diferenciadas, perturbações mínimas causam diversas pequenas avalanches e raras avalanches largas. Assim, grupos locais se comunicam frequentemente por meio de pequenas avalanches e regiões distantes se comunicam menos por avalanches largas. Com esses mecanismos, sistemas à beira do caos conseguem praticar computações extremamente complexas.

O estado líquido parece ter capacidade otimizada para evolução, principalmente pela facilidade de mudança e adaptação. Como Kauffman (1991, p. 84, tradução nossa) explica:

> Assim como ensinou Darwin, mutações e a seleção natural podem aprimorar um sistema biológico pela acumulação de variantes mínimas sucessivas, assim como um técnico pode aprimorar a tecnologia. Apesar disso, nem todos os sistemas possuem a capacidade de se adaptar e aprimorar assim. (...) Redes na fronteira entre ordem e caos podem ter a flexibilidade de se adaptar rapidamente e de forma bem sucedida durante a acumulação de variações úteis. Em tais sistemas equilibrados, a maioria das mutações possuem pequenas consequências por causa da natureza homeostática do sistema. Um grupo de mutações, entretanto, causam correntes de mudança maiores. Sistemas em equilíbrio irão, portanto, tipicamente adaptar-se a um ambiente em mudanças gradualmente, mas se necessário, podem ocasionalmente mudar rapidamente. Essas propriedades são observadas em organismos. Se redes Booleanas de processamento paralelo equilibradas entre a ordem e o caos podem se adaptar mais facilmente, então elas podem ser o alvo inevitável da seleção natural. A habilidade de abusar da seleção natural seria uma das primeiras características selecionadas[3].

Será interessante para a análise que segue notar como algumas propriedades emergentes dos sistemas são apenas possíveis, ou pelo menos, mais facilmente atingíveis em sistemas líquidos com alto grau de auto-organização entre seus elementos. A partir destas noções pretende-se estudar a constituição de uma meta em organismos e metas em máquinas.

4. Metas Genuínas e Metas Parciais

Diante do exposto nas seções anteriores lança-se mão da ideia de microagentes de Dennett junto à teoria de auto-organização na evolução de Kauffman para diferenciar "metas genuínas" de "metas parciais". As primeiras podem ser definidas enquanto metas que envolvem o interesse de todos ou da maioria dos microagentes de um sistema líquido. Por sua vez, metas parciais são metas que não necessariamente envolvem o interesse de todos ou da maioria dos microagentes do sistema e não são, de forma direta, causalmente traçáveis às metas genuínas.

As metas que os organismos desenvolvem durante a vida são variadas, principalmente em seres humanos, nunca é possível conhecer a meta de uma pessoa sem conhecer no mínimo partes de suas intenções. Entretanto, existe uma meta básica em comum para todos os organismos, a busca da sobrevivência. Esta meta praticamente independe da consciência, os indivíduos agem de acordo com ela, de modo que questionar esta meta seria algo quase automaticamente doentio. Esta meta já está intrínseca ao próprio corpo, as células humanas possuem esta meta. Ser vivo é ter a meta egoísta de sobreviver e também de se replicar. Pode ser questionado que nem todos os seres humanos possuem o interesse de replicação, alguns passam toda a vida sem ter filhos. Entretanto, mesmo não seguindo esta meta, ela ainda assim está embutida em seu corpo. A estratégia embutida para realização desta meta é a necessidade da presença de outros (logo, a fuga à solidão), a necessidade de contato físico, a necessidade (para certas pessoas até incontrolável) e o prazer do ato sexual. Estas estratégias fazem parte do corpo de todo organismo que desenvolveu o sexo como garantia de replicação, e funcionam na maior parte dos casos.

Os seres humanos, de certo modo, também fazem uso de metas parciais. Por exemplo,

quando uma pessoa que não tem interesse em preservar o ambiente e não tem interesse em ser correto socialmente encontra um papel usado no chão e automaticamente o joga no lixo. Ela provavelmente está meramente realizando a ação que outros o mandaram fazer, mas essa meta não tem o mesmo poder das metas genuínas. Elas não estão ligadas a direção básica da ação de cada microagente que habita o corpo. Algumas metas são como as de computadores, mas muitas delas são metas derivadas das metas que em algum pequeno grau estão ligadas à noção de sobrevivência e replicação e o interesse de microagentes do corpo humano. A exemplo tem-se uma pessoa jogando um *vídeo game*, esta executa metas para atingir seus objetivos virtuais. As metas são relativamente parciais, uma vez que, apesar de serem virtuais, possuem um vínculo com o prazer, e o prazer está sempre relacionado a metas genuínas. Pode-se chamar estas de metas derivadas, pois são facilmente traçáveis às metas genuínas. Outro exemplo desta seria a realização de um trabalho burocrático para uma pessoa que trabalha com isso, as pequenas metas do trabalho ainda estão ligadas à meta de sobrevivência e replicação, pois na sociedade atual, realizar com eficiência trabalhos burocráticos é um meio para garantir sobrevivência e (a partir, por exemplo, do *status* financeiro) replicação. Para um computador, jogar um *vídeo game* ou exercer trabalhos burocráticos são metas parciais, elas não vão ter uma rede causal que ligue esta meta a outras, intrínsecas à sua própria estrutura física.

Será que inserir a meta de sobrevivência em uma máquina é apenas dar o comando para que ela tome todas as atitudes necessárias visando manter-se fora de perigo? A linha de estudo que este trabalho segue leva a concluir que não, uma vez que para uma máquina ter metas genuínas seria preciso que todas as suas micro partes estivessem ligadas a um interesse intrínseco e egoísta de sobrevivência e replicação, de tal forma que a função de cada micro parte desta máquina tivesse certo impulso a replicação e sobrevivência. Não apenas isso, todos esses interesses precisariam estar de tal modo organizados ao ponto de formarem uma sincronia (líquida), um grupo auto-organizado de microagentes com interesses comuns. Só a partir desses interesses de cada microagente do agente é que se poderiam derivar metas genuínas. Especula-se que a tentativa de aplicar metas a máquinas falha por serem metas incompatíveis com a estrutura física destas. São metas executadas por processos superficiais e não intrínsecos à organização do sistema. As metas não atingem o sistema como um todo, mas apenas meros processos automatizados relacionados a este sistema. Portanto, a meta de uma máquina não é a meta da própria máquina, não é uma meta do "sistema máquina".

A diferença funcional entre sistemas com metas genuínas de sistemas com apenas metas parciais é que os primeiros conseguem utilizar estas para modificar sua própria estrutura para atingir seus objetivos. Já o sistema com apenas metas parciais não consegue modificar sua estrutura para atingir os objetivos impostos, dado que estes objetivos não são do sistema em si, não resultam das relações internas dos elementos em conjunto que geram essas metas. Por essa razão, nunca se verá máquinas tradicionais se esforçando sem parar, usando toda a energia do sistema para realizar suas metas e sofrendo com a "morte". A não realização das metas não gera o fim do sistema. Em contraste, a não realização de metas genuínas gera eventualmente o fim do sistema, e o sistema como um todo luta para que isto não ocorra. Esta pode ser outra forma de identificar sistemas com metas genuínas, o não cumprimento destas tende a gerar o fim do sistema.

Desse modo, aplicar o termo 'sistema' para máquinas é possível apenas em um sentido epistemológico, sendo mais difícil de fazer o mesmo em um sentido ontológico. A construção de um castelo a partir de blocos infantis pode ser chamada, também, de sistema, mas esse termo será um pouco mais fraco, por ter menos propriedades de um sistema do que outro no qual todas as partes se relacionam com um longo histórico de causação em comum e com fortes relações de interdependência, cuja interação faz surgir um produto emergente.

Por esse ponto de vista, uma máquina está mais longe de metas genuínas e consciência do que seres vivos simples como uma bactéria, alga ou estrela-do-mar. Esses três últimos são sistemas líquidos cujos microagentes interiores possuem interesses em comum capazes de guiar a direção e, portanto, de criar metas para o sistema como um todo.

Se esta análise estiver correta, esta pode ser uma dificuldade para a Robótica e Inteligência

Artificial Forte, porque as máquinas, pelo menos por enquanto, não possuem a estrutura necessária para o surgimento de metas genuínas. Como Kauffman mostra, existem diversos tipos de sistemas ordenados, mas é preciso, exatamente, de certo tipo de sistema para que haja vida e, acrescenta-se, metas genuínas. Esse tipo de sistema é o sistema líquido. Quando um ser vivo faz uma cópia de si é importante que ela seja em muitos aspectos semelhante ao original, mas ainda é essencial para a evolução que tal processo não seja rígido; isto porque como já mencionado, as pequenas divergências que podem ocorrer do padrão são exatamente o que gera a mudança e a novidade na vida. Supondo a criação de máquinas robóticas como as de hoje em dia, mas cuidando de embutir nelas a meta parcial de se replicar, se houvesse uma pequena falha sequer no processo, esta não iria funcionar. Isto porque máquinas são sistemas sólidos (demasiadamente ordenados), não líquidos. Os organismos em contrapartida possuem uma liquidez intrínseca. Para um ser vivo não conseguir se replicar precisa haver muitas falhas no processo, e se existirem pequenas falhas, o ser vivo pode nascer diferente, com uma síndrome genética, por exemplo, mas ainda funcionará como ser vivo. Logo, as máquinas como são atualmente não conseguirão ter metas genuínas incorporadas por não serem compostas de microagentes com interesses e não serem líquidas para que sua própria estrutura adquira este interesse.

As metas necessárias para implementar o funcionamento de um Espaço de Trabalho Global em uma máquina, em grande parte, podem ser metas parciais. Mas com apenas isso, não se conseguiria chegar a máquinas verdadeiramente conscientes. Para chegar a máquinas com intenções e consciência ou a uma inteligência artificial, da forma em que são visualizadas na ficção, a máquina precisaria adquirir liquidez para desenvolver micro unidades com "interesses" próprios.

Um crítico poderia objetar que mentes são compostas de padrões informacionais que possuem metas sem precisar ter relação com interesse de microagentes que compõem o corpo dos organismos desta, e que mentes informacionais poderiam ser produzidas em máquinas. Entretanto, mentes não surgiram do nada, evoluíram de sistemas nervosos com sistemas de interesse muito antigos e foram construídas em cima destes antes mesmo de existir qualquer tipo de cérebro, ou mentes humanas. Ainda, o próprio cérebro é composto desses microagentes com "interesses". Esta ideia é sugerida também pelo fato de cada célula carregar em si uma cópia do DNA. Também, o próprio sistema imune carrega em si metas genuínas, que são do interesse de todo o sistema e precisam diferenciar entre o mundo externo e o agente, sendo fundamental para o início da noção interna de um agente. Como afirma Dennett (1991), quando se tem o objetivo de autorreplicação, criar fronteiras é importante. O organismo não pode ter o objetivo de preservar todo o mundo, ele precisa preservar apenas a si, tornando-se egoísta. Esta é a marca não só da vida, mas do início da agência.

5. Considerações Finais

O filósofo Daniel Dennett, cujas ideias serviram de base para este trabalho, provavelmente não concordaria com a conclusão a que este chegou. A intensidade de sua crítica seria ainda mais sensível se a ideia aqui descrita não trouxesse alguma utilidade para a ciência ou para a compreensão do tópico. Há de se concordar com o autor neste ponto, dizer que máquinas nunca terão metas genuínas não terá efeito prático algum. Mesmo se o objetivo fosse evitar a criação de uma inteligência artificial por algum motivo ético, esse tipo de especulação não seria o bastante para convencer cientistas e engenheiros a abandonarem seus projetos. E nem deveria ser, pois mesmo isto sendo verdade, verificar na prática o que poderia ser feito com o conhecimento alcançado de fato seria mais interessante. Logo, é necessário que haja algum tipo de contribuição do comentário deste trabalho.

Acredita-se, portanto, que esta leitura das metas genuínas pode ressaltar a importância de máquinas que se repliquem para se tornarem verdadeiros agentes, para que processos de seleção desenvolvam micro partes que criem seus próprios "interesses". Existem duas vertentes que podem algum dia promover este tipo de inteligência artificial, a vertente da nanorrobótica (Sierra, Weir & Jones, 2005) e a das máquinas híbridas,

misturas de células vivas e máquinas (Reger, Fleming, Sanguineti, Alford & Mussa-Ivaldi, 2000).

A partir da primeira vertente podem surgir sistemas compostos de nanoagentes capazes de produzir liquidez ao sistema. Com a segunda, as metas da parte mecânica podem ser derivadas das metas genuínas das células vivas da parte biológica. Pode-se pensar neste caso como dois sistemas funcionando em conjunto, o sistema biológico possui as metas genuínas como outros seres vivos e o sistema mecânico obtém do sistema biológico suas metas derivadas. Logo, acredita-se que os ideais da Inteligência Artificial Forte e a Robótica podem chegar a vários resultados efetivos se unidos a novas ciências e métodos.

Destarte, conclui-se que para a implementação de aspectos tão complexos como funções conscientes do Espaço de Trabalho Global, seria preciso que houvesse um grupo auto-organizado de agentes com "interesses" para executá-los. A necessidade de implementar um Espaço de Trabalho Global precisa ser uma necessidade do sistema em si, não de um agente externo ao sistema. A única forma de forçar o surgimento de um Espaço de Trabalho Global em um sistema de forma externa é imprimindo pressão seletiva em sistemas líquidos. As metas parciais podem chegar a vários objetivos, mas não ao objetivo de construir um agente semelhante ao vivo, por faltar metas como as de sobrevivência e replicação, as quais são intrínsecas ao funcionamento do corpo do agente situado no mundo.

6. Agradecimentos

Agradecemos à Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento (bolsa CAPES).

Referencias Bibliográficas

Baars, B. J. (1988). *A Cognitive Theory Of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. J., & Franklin, S. (2007). An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. Neural Networks, v.20, 955-961.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1986). *The Blind Watchmaker*. New York: Norton.

Debrun, M. (1996). A idéia de Auto-Organização. In: Debrun, M., Gonzalez, M. E. Q., Pessoa Jr, O. (orgs.). Campinas: Auto-Organização: estudos interdisciplinares. v.18, 3-23.

Dennett, D. C. (1991). *Consciousness Explained*. New York: Black Bay Books.

Dennett, D. C. (1995). Darwin's Dangerous Idea: evolution and the meanings of life. New York: Penguin Books.

Franklin, S. & Patterson Jr, F. (2006), The LIDA architecture: adding new modes of learning to an inteligent, autonomous, software agent. Integrated Design and Process Technology, IDPT-2006, june, 1-7.

Kauffman, S. (1991). Antichaos and adaptation. *Scientific American*, ago, 78-84.

Kauffman, S. (1993). *The origins of order*. New York: Oxford University Press.

Madl, T., Baars, B. J. & Franklin, S. (2011). The timing of the Cognitive Cycle. *Plos one*, v.6 n.4, 1-16.

Maslow, A. (1970). *Motivation and Personality*. 2ed. New York: Harper & Row.

Pereira Jr., A., Paleari, L., Costa, F., & Guimarães, R. (2004). Evolução Biológica e Auto-Organização: apresentando, discutindo e exemplificando uma proposta teórica. In: D'Ottaviano, I. M. L., Gonzales, M. E. Q. & Souza, G. M. (orgs.). Campinas: Auto-Organização: estudos interdisciplinares, v.39.

Pereira Jr., A. & Pereira, M. (2010) Teoria da Auto-Organização: uma Introdução e Possível Aplicação nas Ciências da Saúde. Revista Simbio-Logias, v.3, n.5, 102-114.

Reger B. D., Fleming K. M., Sanguineti V., Alford S. & Mussa-Ivaldi F. (2000). Connecting brains to robots: an artificial body for studying the computational properties of neural tissues. *Artificial Life*, v.6, 307–324

Sierra, D. P., Weir, N. A. & Jones, J. F. (2005). A review of research in the field of nanorobotics. U.S. Department of Energy - Office of Scientific and Technical Information, Oak Ridge.

Silva, R., Gudwin, R. (2011). Developing a Consciousness-Based Mind for an Artificial Creature. *Lecture Notes in Computer Science*, v.6404, p.122-132.

Notas

(1) One helpful step to solve this communication problem is to make public a global message on a large blackboard in front of the auditorium, so that in principle anyone can read the message and react. In fact, it would only be read by experts who could understand it or parts of it, but one cannot know ahead of time who those experts are, so that it is necessary to make it potentially available to anyone in the audience. At any time a number of experts may be trying to broadcast global messages, but the blackboard cannot accommodate all of the messages at the same time --- different messages will often be mutually contradictory. So some of the experts may compete for access to the blackboard, and some of them may be cooperating in an effort to broadcast a global message.

(2) Vale notar que esta abordagem do contexto enquanto representação se mostra ingênua em relação ao problema dos *Frames*. Este é um dos maiores problemas da ciência cognitiva e foi primeiro identificado no processo de tentar representar o conhecimento necessário para ações, mais especificamente, ao modelar formalmente as mudanças e as não mudanças de um ambiente.

(3) As Darwin taught, mutations and natural selection can improve a biological system through the accumulation of successive minor variants, just as tinkering can improve technology. Yet not all systems have the capacity to adapt and improve in that way [...]. Networks on the boundary between order and chaos may have the flexibility to adapt rapidly and successfully through the accumulation of useful variations. In such poised systems, most mutations have small consequences because of the systems' homeostatic nature. A few mutations, however, cause larger cascades of change. Poised systems will therefore typically adapt to a changing environment gradually, but if necessary, they can occasionally change rapidly. These properties are observed in organisms. If parallel-processing Boolean networks poised between order and chaos can adapt most readily, then they may be the Inevitable target of natural selection. The ability to take advantage of natural selection would be one of the first traits selected.