

**THE WEB OF BELIEF:  
UMA PERSPECTIVA DA INTELIGÊNCIA ARTIFICIAL<sup>1</sup>**

João Cachopo e Ana Cardoso

*I would never die for my beliefs because I might be wrong.*  
Bertrand Russell

### **1. Introdução**

Algumas das características que nos distinguem, como seres humanos, dos outros animais são a nossa capacidade de raciocinar, de compreender e utilizar de forma eficaz a informação que obtemos do ambiente onde estamos inseridos, e de construir planos para atingir objectivos.

Este tipo de comportamento só é possível se mantivermos uma representação mental do ambiente que nos rodeia, não só no que diz respeito ao estado em que se encontra, mas também ao modo como evolui. Cada um de nós mantém, assim, um conjunto de *crenças*<sup>2</sup> acerca do que se passa à nossa volta.

Naturalmente, dada a complexidade do mundo real e as limitações inerentes aos nossos mecanismos de percepção, estas crenças constituem apenas um modelo aproximado da realidade. Em muitos casos, o estado do ambiente que nos rodeia é muito diferente daquilo que nós *acreditamos* ser, o que acontece não só por termos informação incompleta sobre o ambiente mas também pelo facto de o próprio ambiente ser dinâmico, isto é, estar em mudança. Por este motivo, é evidente que o nosso conjunto de crenças não

<sup>1</sup> Gostaríamos de agradecer ao Professor Doutor João Branquinho pelas conversas que manteve connosco sobre estes temas e pelos comentários e sugestões feitos a uma versão preliminar deste artigo.

<sup>2</sup> O termo «crença» tem dois significados possíveis: (1) algo que é acreditado e que pode ser verdadeiro ou falso, ou seja, uma proposição; ou (2) o acto de acreditar numa proposição. Neste artigo, o termo «crença» é usado com o significado dado em (1). No entanto, a existência de entidades mentais desta natureza não é aceite por todos.

se mantém imutável. Muito pelo contrário, está permanentemente a sofrer alterações à medida que obtemos nova informação: novas crenças são formadas e crenças antigas podem ter de ser abandonadas. A forma como estas alterações se processam depende não só da informação que obtemos do mundo exterior, mas também da estrutura das nossas próprias crenças: das dependências existentes entre elas, da sua origem, do grau de confiança que temos nelas, etc.

No livro *The Web of Belief*, Willard Quine e Stephen Ullian (Quine e Ullian 1978) fazem um estudo das crenças, dando particular ênfase aos seus aspectos dinâmicos, isto é, à forma como as mudanças nas crenças ocorrem.

Neste artigo propomo-nos fazer um resumo e uma apreciação do contributo dado por este livro, mas de um ponto de vista diferente do da filosofia tradicional: o da inteligência artificial (IA). Para motivarmos esta nossa abordagem, na próxima secção descrevemos em que medida é que o estudo das crenças e dos seus aspectos dinâmicos é relevante para a IA, dando uma perspectiva das principais preocupações existentes nesta área. Nas restantes secções fazemos um resumo dos aspectos mais importantes do livro, tendo em conta essa perspectiva.

## 2. Crenças e mudança de crenças em IA

### 2.1. Os objectivos da IA

A IA é uma área interdisciplinar que congrega contribuições de várias áreas, como a filosofia, a ciência da computação, a psicologia, a sociologia, etc.

O principal objectivo da IA foi, durante muitos anos, construir uma máquina que apresentasse comportamento inteligente idêntico ao do ser humano.<sup>3</sup> Para atingir este objectivo é necessário, naturalmente, *compreender* o comportamento inteligente e conseguir modelá-lo numa máquina. Este objectivo só por si já evidencia claramente a importância do estudo das crenças para a IA. Como vimos na introdução, as actividades que distinguem

<sup>3</sup> Este objectivo é contestado por alguns críticos da IA, que afirmam ser impossível existir inteligência dissociada do ser humano, isto é, numa máquina. Um dos opositores mais veementes à IA é o filósofo John Searle, autor do famoso argumento do «quarto chinês» (Searle 1980, 1982, 1990).

o ser humano como ser inteligente pressupõem a existência de uma representação do mundo que nos rodeia, o que corresponde ao papel que as crenças desempenham. Da mesma forma, para podermos construir uma máquina que consiga raciocinar, compreender e utilizar de forma eficaz a informação que recebe, e ainda estabelecer planos para atingir objectivos, é necessário manter nessa máquina um conjunto de *crenças* que suportem estas actividades.

No entanto, com o amadurecimento da área, o objectivo da IA para a maior parte dos seus investigadores tem vindo a mudar. No nosso entender, isto deve-se fundamentalmente a dois problemas que o objectivo anterior apresentava: por um lado, é extremamente difícil caracterizar o que é a *inteligência*; por outro lado, de um ponto de vista tecnológico, não faz muito sentido tentar simular um ser humano se isso significar construir uma máquina que cometa os mesmos erros que um ser humano comete. Assim, ao invés de tentar construir uma máquina com comportamento inteligente (leia-se, semelhante ao de um ser humano), a IA moderna centra os seus objectivos na construção de uma máquina com comportamento *racional*.

O critério de racionalidade, na verdade, é mais exigente que o anterior, mas também mais simples de caracterizar e mais útil do ponto de vista tecnológico. Isto permite definir o objectivo da IA com mais rigor: construir sistemas automáticos (máquinas) que respondam sempre da melhor forma possível aos vários problemas que tenham que resolver. Este objectivo afasta-nos um pouco do comportamento de um ser humano, uma vez que os seres humanos nem sempre apresentam comportamento racional: tomam decisões erradas, cometem erros de raciocínio, etc. Apesar de este livro considerar as crenças do ponto de vista do ser humano, torna-se particularmente relevante para a IA precisamente pelo facto de se preocupar essencialmente com o estudo das crenças racionais.

O tipo de raciocínio que Quine e Ullian consideram para basear o seu estudo das crenças é o raciocínio científico. Segundo eles, a ciência é fruto da investigação racional, da aplicação criteriosa da lógica ao conhecimento adquirido.

Na introdução do livro os autores criticam fortemente as doutrinas anti-racionalistas e apontam algumas das razões para a sua existência, embora as refutem completamente. O seu objectivo neste livro, como eles próprios afirmam, é identificar critérios que permitam distinguir as crenças que são razoáveis das que não o são.

## 2.2 O papel das crenças na construção de um agente

Em IA, tal como na área da filosofia da mente, utiliza-se o termo *agente* com o significado de algo que está inserido num ambiente, com capacidade para perceber e efectuar acções nesse ambiente. Segundo esta definição, o objectivo da IA é construir um *agente racional*. Nesta secção vamos ver qual o papel que as crenças desempenham na obtenção desse objectivo. Ao analisarmos as principais dificuldades que surgem na construção de um agente, compreendemos melhor em que aspectos o trabalho de Quine e Ullian contribui para o avanço da IA.

Um dos componentes fundamentais para a construção de um agente racional é o conhecimento. O conhecimento é essencial para que o agente compreenda o que se passa à sua volta, raciocine acerca das propriedades dos objectos, possa prever a evolução do mundo, efectue planos de modo a atingir objectivos, etc.<sup>4</sup>

No livro e no resto deste artigo considera-se que o conhecimento de um agente é representado por um conjunto de crenças — proposições acerca do mundo que o agente julga serem verdade — independentemente da forma como esse conjunto de crenças poderia ser representado numa máquina.

Para além dos problemas relacionados com a representação do conhecimento, que corresponde a uma das sub-áreas da IA, existem muitos outros problemas relacionados com o conhecimento, como a questão de saber como pode o agente adquirir conhecimento na sua interacção com o mundo, isto é, *aprender* autonomamente, e como pode o agente utilizar o seu conhecimento para *raciocinar*, *planear*, *compreender* ou *actuar* sobre o mundo. Muitas destas actividades provocam alterações no próprio conhecimento que o agente mantém.

A sub-área da IA que se preocupa com a forma como essas alterações se processam é a área de *revisão de crenças*, que estuda as dependências existentes entre as várias crenças de um agente e o que fazer quando essas crenças sofrem alterações. Por exemplo, quando o agente recebe nova informa-

<sup>4</sup> Embora existam algumas abordagens em IA que explicam parte do comportamento de um agente com mecanismos mais primitivos, baseados em reflexos, as actividades racionais mais complexas, como o planeamento, por exemplo, exigem a utilização de conhecimento por parte do agente.

ção que entra em contradição com a que já mantinha é necessário, para que o agente tenha um comportamento racional, resolver essa contradição, abandonando algumas das suas crenças. Infelizmente, em geral, não é fácil determinar que crenças devem ser abandonadas, uma vez que podem existir várias alternativas, cuja escolha depende de muitos factores, também eles muitas vezes difíceis de determinar.

Se considerarmos que a própria definição de agente pressupõe que este está inserido num ambiente com o qual está em constante interacção, é evidente que a compreensão dos processos que determinam a forma como as crenças devem ser alteradas é de vital importância para a construção de um agente.

A forma como Quine e Ullian começam o segundo capítulo do livro *The Web of Belief* torna evidente a preocupação fundamental dos autores neste livro — os aspectos dinâmicos das crenças:

One's repertoire of beliefs changes in nearly every waking moment. The merest chirp of a bird or chug of a passing motor, when recognized as such, adds a belief to our fluctuating store. (Quine e Ullian 1978, 9)

Neste livro, discute-se como devem as mudanças processar-se e quais os critérios que determinam que crenças abandonar e que crenças manter quando se verifica uma mudança. Para isso, consideram-se alguns dos aspectos mais importantes das crenças, como por exemplo, a sua origem e as suas propriedades.

### 3. A teia de crenças

A referência nas secções anteriores ao conjunto de crenças de um agente pode dar a ideia de que as crenças de um agente são entidades separadas, sem relações entre si para além do facto de todas elas serem crenças. Neste aspecto, o título do livro é particularmente sugestivo, evidenciando claramente que os autores adoptam a posição holista, defendendo a existência de uma estrutura subjacente às crenças de um agente. Isto corresponde também à posição adoptada pela IA, que considera que as crenças mantêm relações entre si, especialmente relações inferenciais, criando uma estrutura onde todas as crenças podem estar directa ou indirectamente interligadas. Por exemplo, existe certamente uma relação entre a crença de que Sócrates era

mortal e as crenças de que todos os homens são mortais e de que Sócrates era um homem. Dizemos que as duas últimas *suportam* a primeira.

O que torna o processo de mudança das crenças mais complicado é precisamente a existência destas relações entre as crenças. Não basta adicionar ou retirar uma crença ao «conjunto» de crenças; é preciso também ter em linha de conta quais as relações existentes entre a crença a adicionar ou a retirar e todas as outras crenças. O exemplo paradigmático é o da adição de uma crença que entre em contradição com outra anteriormente mantida pelo agente. Por exemplo, que deverá acontecer se forem obtidos dados de que de facto Sócrates não era mortal?

Racionalmente, não faz sentido manter crenças (que se sabe serem) contraditórias. Assim sendo, sempre que uma situação destas ocorre é necessário abandonar alguma das crenças em conflito. Repare-se, no entanto, que não é solução abandonar apenas a crença de que Sócrates é mortal, uma vez que esta crença é uma consequência lógica das crenças que a suportam, nomeadamente de que todos os homens são mortais e de que Sócrates era um homem; logo, pelo menos uma destas deve ser abandonada.

A resolução de conflitos entre crenças contraditórias é o tipo mais problemático de alteração nas crenças, mas também o mais interessante. A alteração verificada nas crenças de um agente como resultado da resolução de um destes conflitos é designada por *revisão*, e é este o tipo de mudança mais discutido no livro.

O segundo capítulo dedica-se inteiramente ao problema da mudança das crenças e, em particular, à revisão das crenças. Para além disso, faz ainda algumas distinções importantes no contexto da IA, como por exemplo, a diferença entre crença e conhecimento. Este capítulo, apesar de pequeno, constitui, no nosso entender, e para os nossos propósitos, o núcleo de todo o livro.

Nos capítulos três a seis são considerados vários tipos de crenças: as que resultam da *observação* e do *testemunho*, bem como as crenças *auto-evidentes* e as *hipóteses*. Estas crenças constituem a base de todas as outras crenças, mas não são, elas próprias, logicamente suportadas por outras crenças. A separação entre crenças básicas e crenças derivadas, que dependem logicamente das crenças básicas, revela a atitude fundacionalista tomada pelos autores, e que está subjacente a todo o trabalho de revisão de crenças efectuado em IA. A distinção entre os vários tipos de crenças básicas é

fundamental para o processo de revisão, uma vez que a importância de uma crença neste processo depende da sua origem.

Destes vários tipos de crenças, as hipóteses são as mais interessantes e constituem o tema dos restantes capítulos do livro (do capítulo sete em diante), que se dedicam a explorar vários aspectos relacionados com este tipo de crenças. Por este motivo, relativamente aos capítulos três a cinco, vamos apenas referir, muito sumariamente, os seus aspectos mais importantes, mas sem entrar em muitos detalhes, concentrando a nossa atenção nos últimos capítulos do livro, que dizem respeito às hipóteses.

#### **4. A mudança de crenças**

O conjunto de crenças de um agente sofre alterações sempre que ele tem que incorporar nova informação nas suas crenças, quer essa informação tenha vindo do exterior, quer seja resultado do seu próprio processo de raciocínio.

No entanto, as mudanças nas crenças de um agente não consistem apenas na adição de novas crenças. Por vezes, algumas crenças são também abandonadas.

No livro, os autores consideram que existem fundamentalmente duas situações nas quais as crenças podem ser abandonadas: por serem pouco importantes ou por entrarem em conflito com outras crenças. A primeira situação não é discutida no livro e, mesmo do ponto de vista da IA, não é normalmente considerada. Embora possa ser muito importante para compreendermos o modo de funcionamento da memória humana, este aspecto pode ser relativamente ignorado no estudo de um agente racional. Por isso, apenas é considerada a segunda situação, a revisão de crenças, que é também a mais importante para a IA.

A decisão de qual deve ser a crença a abandonar numa revisão não é simples e baseia-se numa série de critérios relacionados com as características das crenças e das relações entre elas. Segundo Quine e Ullian um agente racional deve escolher as crenças a abandonar com base nos *dados*<sup>5</sup> existentes para cada uma delas. Para essa decisão são utilizadas crenças de ordem superior, que entram em linha de conta com esses dados.

<sup>5</sup> Os dados para uma crença são as informações usadas para a justificar.

Embora por vezes as *causas* para a formação de uma determinada crença sirvam como dados que suportam essa crença, isto nem sempre acontece. Por isso, é importante distinguir entre as causas e os dados para uma crença. Enquanto uma determinada crença não entrar em conflito com as restantes, é razoável aceitá-la, qualquer que seja a sua causa, mesmo que não tenhamos dados que a suportem. No entanto, assim que ela entrar em algum conflito, temos que considerar os dados que temos para ela, nos quais vamos basear a nossa escolha.

Tal como as causas para uma crença, a *intensidade* com que se mantém uma crença não pode ser confundida com os dados que a suportam, embora num agente racional a intensidade com que uma crença é mantida deva estar directamente relacionada com os dados existentes para essa crença. Do ponto de vista puramente lógico, se existe uma contradição entre as crenças de um agente é porque uma dessas crenças tem de ser necessariamente falsa. Infelizmente, no processo de revisão, nem sempre é possível determinar com base apenas em critérios racionais, que crença é essa. Neste caso, a única alternativa é escolher algumas crenças com menos dados e abandoná-las, não acreditando também na sua negação. Neste caso temos uma *revisão não decisiva*. Por outro lado, quando conseguimos identificar qual a crença que não pode ser verdadeira, podemos abandoná-la e passar a acreditar na sua negação. Neste caso temos uma *revisão decisiva*.

Como veremos, um dos factores mais importantes a ter em linha de conta na escolha das crenças a abandonar em caso de conflito é a origem das crenças ou dos dados que as suportam. Na próxima secção vamos apresentar os vários tipos de crenças considerados no livro.

### 5. Os vários tipos de crenças

Num agente racional ideal<sup>6</sup> a única forma de resolver um conflito é abandonar algumas das crenças básicas, que não dependem, para a sua justificação, de nenhuma outras. De acordo com a teoria fundacionalista, as crenças derivadas são baseadas única e exclusivamente em crenças básicas. Por este motivo, o abandono de uma crença derivada não permite resolver o conflito, uma vez que, sendo o agente logicamente omnisciente, a crença

<sup>6</sup> Considera-se um agente racional ideal um agente que seja logicamente omnisciente, isto é, que acredite em todas as consequências lógicas daquilo em que acredita.



derivada abandonada voltaria a ser acreditada através do raciocínio lógico do agente. Mesmo considerando um agente racional real (com uma capacidade limitada de raciocínio) continua a ser preferível abandonar crenças básicas, uma vez que nos comprometemos com a teoria fundacionalista.

Nesta secção vamos falar de crenças básicas: as que resultam da *observação*, as crenças *auto-evidentes*, as que resultam da interacção com outros agentes (*testemunhos*) e as *hipóteses*.

### 5.1. Observação

As primeiras crenças consideradas no livro são as que resultam da *observação*. Estas crenças são fruto da percepção que o agente tem do mundo que o rodeia.

Os dados existentes para este tipo de crenças — o facto de resultarem da observação — fazem com que normalmente não sejam questionáveis. Um agente, quando confrontado com a necessidade de abandonar alguma crença, coloca grandes entraves ao abandono deste tipo de crenças. Uma das poucas situações em que uma crença resultante de uma observação é posta em causa é quando vai contra uma teoria<sup>7</sup> que já previu correctamente muitos acontecimentos e que já acomodou muitos fenómenos relevantes. No entanto, mesmo neste caso, quando essa teoria começa a forçar o abandono de muitas crenças resultantes da observação, isso constitui um forte sinal de que está errada e de que precisa de ser reformulada para cobrir os casos observados.

Este tipo de crenças constitui, portanto, o núcleo de crenças mais forte de um agente.

### 5.2. Auto-evidência

Tal como as anteriores, as crenças auto-evidentes são crenças que não dependem de nenhuma outra crença. No entanto, não são fruto da observação do mundo exterior.

No livro, onde as crenças estão intimamente relacionadas com as frases utilizadas para as representar, as crenças auto-evidentes são apresentadas como dependendo apenas, quanto à sua verdade, do significado das palavras que constituem cada frase. Nestas crenças estão incluídas as definições (por

<sup>7</sup> Uma teoria é um conjunto articulado de crenças.

exemplo, que os solteiros não são casados) e as tautologias (verdades lógicas), que não podem deixar de ser verdadeiras. É este último tipo de crenças que normalmente é considerado em IA como auto-evidente.

As crenças deste tipo, apesar de poderem entrar em conflito com outras, não podem nunca ser consideradas culpadas da contradição e serem abandonadas por causa disso, uma vez que a sua verdade não pode ser contestada.

### *5.3. Testemunho*

Tal como as crenças que resultam da observação por parte do agente, os testemunhos também têm origem no exterior. No entanto, esta informação resulta da comunicação do agente com outros agentes.

Embora a comunicação entre dois agentes seja feita com a ajuda dos mecanismos de percepção que eles têm para o efeito, as crenças obtidas desta forma não são consideradas como observações. A percepção, neste caso, é utilizada não para perceber directamente a realidade que dá origem à crença, mas sim o relato da percepção realizada por outros agentes. Naturalmente, a credibilidade das crenças obtidas desta forma é radicalmente diferente da obtida através da observação, dependendo da fonte a partir da qual essa informação é obtida.

Ao discutir este problema, Quine e Ullian referem a necessidade de se registar com cada crença obtida por este processo a fonte que lhe deu origem. Esta informação é vital para o processo de revisão, uma vez que a credibilidade da fonte é um dos factores mais importantes para guiar a escolha das crenças a abandonar. Para além disso, sempre que existe alguma alteração na credibilidade de uma fonte podemos rever as crenças que tiveram origem nessa fonte, podendo, por exemplo, abandonar algumas crenças cujos dados tenham sido enfraquecidos.

### *5.4. Hipóteses*

As hipóteses são o último tipo de crenças tratado no livro, mas também o que é tratado com maior profundidade. O seguinte excerto do livro dá uma ideia do papel das hipóteses no conjunto de crenças de um agente:

People adopt or entertain a hypothesis because it would explain, if it were true, some things that they already believe. (Quine e Ullian 1978, 66)

A ideia subjacente a este excerto, é a de que as observações e as crenças auto-evidentes não são suficientes para construir teorias que permitam explicar as observações efectuadas e fazer previsões. As hipóteses são usadas neste caso para completar essas teorias.

A determinação de quais as hipóteses a formar, no entanto, não é trivial. Apesar de existirem vários processos pelos quais as hipóteses podem ser formadas (que são descritos nos restantes capítulos do livro), existem alguns critérios gerais que permitem avaliar as hipóteses na sua adequação a uma determinada teoria. Segundo os autores, existem cinco virtudes que as hipóteses têm em maior ou menor grau:

**Conservadorismo:** quando se introduz uma hipótese, ela deve ser o mais possível compatível com as crenças anteriores, isto é, entrar em conflito com o menor número possível das restantes crenças.

**Modéstia:** devem-se preferir as hipóteses que partem do princípio que as mudanças no mundo foram as menores possíveis e/ou as mais habituais.

**Simplicidade:** esta virtude pode entrar em conflito com as anteriores, mas por vezes vale a pena sacrificar a simplicidade das partes para aumentar a simplicidade do todo.

**Generalidade:** pode colidir com a modéstia; mas quanto mais geral for uma hipótese, mais importante ela se torna se for verdadeira e menos hipóteses são necessárias para explicar os mesmos factos.

**Refutabilidade:** serve para garantir que as hipóteses são confirmadas ou refutadas por observações e desta forma fazem a sua ligação com o mundo exterior. Se não puderem ser testadas, para ver se as suas previsões estão ou não correctas, as hipóteses dificilmente dirão algo de novo.

Um outro aspecto importante acerca das hipóteses é a forma como estas são tratadas no processo de revisão de crenças. Como vimos, as crenças que são formadas a partir da observação raramente são abandonadas por um agente racional. O único caso referido onde isso podia acontecer era quando essas crenças iam contra as previsões de uma teoria já com muito crédito,

isto é, com muitas previsões correctas feitas anteriormente. Uma vez que o principal papel das hipóteses é na construção de teorias, isto dá-nos uma indicação do que utilizar como evidência que sirva de suporte a uma hipótese: é tanto mais difícil abandonar uma hipótese quanto mais sucesso ela tenha tido nas suas previsões.

## 6. Indução, Analogia e Intuição

No capítulo sete, Quine e Ullian discutem dois processos, muito relacionados, que permitem a formação de novas hipóteses: a *indução* e a *analogia*.

A analogia é o processo que permite a um agente prever o que vai acontecer numa determinada situação a partir de situações semelhantes pelas quais tenha passado anteriormente.

A indução leva este processo um pouco mais longe: através da indução, o agente forma uma crença (hipótese) que generaliza a partir de uma série de exemplos semelhantes observados anteriormente. Segundo esta nova hipótese, todas as situações idênticas às observadas têm um grande número de propriedades em comum.

Com isto conseguimos perceber melhor em que sentido é que as hipóteses são usadas para criar teorias. Através do processo de indução são estabelecidas «leis» que descrevem como funciona uma certa parte do mundo. A partir destas hipóteses, o agente consegue prever o que pode acontecer em situações futuras. As hipóteses formadas através deste processo devem satisfazer as virtudes de simplicidade e generalidade descritas na secção anterior.

Quer a indução quer a analogia, são tema de estudo de algumas sub-áreas da IA. Em contrapartida, a *intuição* de que os autores falam no fim do capítulo sete não é considerada em IA. Isto está de acordo com a opinião dos autores, segundo a qual a intuição não tem lugar no processo de raciocínio de um agente racional.

## 7. Confirmação e refutação de hipóteses

As hipóteses são, em grande parte, o resultado de um processo de «adivinhação» por parte do agente. Por isso, é fundamental testar as hipóteses, para validar os resultados obtidos com esse processo.

As virtudes que as hipóteses devem satisfazer são alguns dos critérios que permitem encaminhar este processo de «adivinhação» para a formação de hipóteses plausíveis. No entanto, as virtudes, só por si, não permitem confirmar ou refutar hipóteses.

A forma como se podem testar as hipóteses é o tema do capítulo oito do livro. Segundo os autores, o teste das hipóteses é feito essencialmente através do teste empírico das suas consequências. Para testar uma hipótese é necessário efectuar experiências, verificando se os resultados obtidos estão de acordo com as previsões feitas pela hipótese. Quando isto acontece, obtêm-se mais dados que confirmam a hipótese, aumentando assim a sua credibilidade. No entanto, isso pode ter acontecido por pura coincidência. Para minimizar este efeito, os autores introduzem uma nova virtude que permite caracterizar as hipóteses: a *precisão*. Quanto mais precisa for uma hipótese, mais difícil se torna a ocorrência destas coincidências e mais fácil é caracterizar as experiências necessárias para a testar.

Quando os resultados obtidos numa experiência não estão de acordo com as previsões da hipótese, ela é refutada. No entanto, para a realização de uma experiência, para além da hipótese que está a ser testada, é normalmente necessário supor que algumas condições se verificam. Por isso, o resultado negativo de uma experiência nem sempre permite a refutação de uma hipótese, uma vez que a responsabilidade pelo erro da previsão pode caber à falsidade de uma das suposições adicionais.

Para terminar, é importante referir o facto de que as hipóteses utilizadas no dia a dia (que em termos de IA correspondem ao conhecimento de senso comum) têm normalmente excepções. Na nossa opinião, mais vale ter hipóteses simples e gerais que por vezes não prevêm correctamente os acontecimentos, do que ter hipóteses infalíveis mas que são demasiado complexas e/ou pouco gerais.

## **8. Explicação, persuasão e avaliação**

Os dois últimos capítulos do livro tratam de assuntos que, na nossa opinião, são pouco importantes do ponto de vista da revisão de crenças, embora tenham alguma relevância para a compreensão da estrutura das crenças.

O capítulo nove aborda a necessidade inata do ser humano de encontrar explicações para as suas crenças: não só para os fenómenos observados, mas também para as hipóteses formuladas para os explicar. As explicações exis-

tentes para as hipóteses são importantes porque a credibilidade de uma explicação é transmitida como um dado adicional para as hipóteses que ela explica. Adicionalmente, a procura da explicação para uma hipótese é outra forma de a confirmar.

No capítulo dez é tratado o tema da persuasão e da avaliação das crenças, isto é, da utilização da capacidade de comunicação de um agente para transmitir ou inculcar crenças noutros agentes. A linguagem, segundo os autores, pode ser usada para dois fins distintos: aprender (adquirir crenças) com outros agentes e convencê-los a efectuar certas acções. Os dados que um agente conseguir reunir como suporte para uma determinada crença são fundamentais para conseguir persuadir outros agentes no sentido de virem a formar essa crença.

### 9. Apreciação global

*The Web of Belief* trata de um assunto extremamente interessante e particularmente relevante para a área da inteligência artificial: a revisão de crenças.

O livro está escrito de uma forma bastante acessível, sem exigir muitos conhecimentos prévios de filosofia, o que facilitou a sua análise do nosso ponto de vista.

O facto de os autores se preocuparem com o estudo do pensamento científico é com certeza um factor muito importante para a sua relevância para a IA, uma vez que ambos se preocupam com o raciocínio de agentes racionais.

Embora no livro a discussão do tema seja feita de uma forma muito informal, difícil de ser utilizada computacionalmente (em particular, no que diz respeito às virtudes das hipóteses), muitas das ideias apresentadas podem ser aproveitadas na construção das teorias de revisão de crenças necessárias em IA. Do nosso ponto de vista, os capítulos mais importantes são os capítulos dois a seis inclusive, onde se descreve o problema da revisão de crenças e a natureza dos vários tipos de crenças de um agente. Neste sentido, os capítulos sete e oito são também particularmente relevantes, tendo até relações com outras áreas da IA, como por exemplo a aprendizagem, no caso da indução.

João Cachopo (jcatchopo@gia.ist.utl.pt)  
 Ana Cardoso (acardoso@gia.ist.utl.pt)  
 Sec. Sistemas — DEM  
 Instituto Superior Técnico  
 Av. Rovisco Pais, 1, 1096 Lisboa Codex

*Referências*

- Quine, W. V. e Ullian, J. S. 1978 *The Web of Belief*. 2.<sup>a</sup> edição. Nova Iorque: Random House.
- Searle, John 1980 Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3, 417-424.
- Searle, John 1982 The Chinese Room Revisited: Response to Further Commentaries on «Minds, Brains, and Programs». *Behavioral and Brain Sciences*, 5, 345-348.
- Searle, John 1990 Is the Brain's Mind a Computer Program? *Scientific American*, 262 (1), 20-25.