

INSTITUIÇÃO

Porfírio Silva* e Pedro U. Lima

Instituto de Sistemas e Robótica, Instituto Superior Técnico

As ciências do artificial, designadamente a Inteligência Artificial, experimentaram repetidamente a possibilidade de algumas máquinas serem vistas como parceiras de conversa dos humanos. Se alguns encaram esse projecto como um fracasso constatado, talvez seja interessante compará-lo com outra aposta, seguramente vista como muito mais banal, de... conversar com um humano. Por que encaramos de formas tão diferentes essas duas possibilidades?

No fundo, talvez continuemos surpreendidos pela forma como as máquinas estão entre nós. Pode dar-se o caso de isso se dever a uma certa prioridade que atribuímos à consciência constituente, bem como a uma certa desatenção ao mundo cuja dinâmica fundamental é a instituição. Para compreender essa dinâmica entre o instituído e o a instituir é particularmente relevante considerar o papel da linguagem e da memória.

É o que aqui faremos a partir das nossas experiências em Sociedades Artificiais, onde estamos a tentar abrir uma nova abordagem à noção de instituição, para humanos e máquinas, recorrendo a sistemas de múltiplos *robots*.

Podemos falar com uma máquina?

Quando Alan Turing (1950), para expressar operacionalmente a questão «as máquinas pensam?», propôs o «jogo da imitação», tornou a questão da linguagem central nas ciências do artificial. No que depois se designaria por «teste de Turing», temos numa sala uma pessoa e um computador e noutra sala, separada, outro humano a interrogar aqueles dois e, comunicando apenas por telétipo para que as diferenças físicas não constituam qualquer pista, o interrogador humano tentará, apenas com base no conteúdo das respostas, distinguir o humano da máquina, estando o computador programado precisamente para evitar essa identificação. Uma vez que o computador teria de ser melhorado em termos de capacidade de memória, velocidade de processamento e programação, Turing prevê que por volta do ano 2000 haverá computadores que jogarão tão bem o jogo da imitação que um interrogador hu-

* A investigação do primeiro autor beneficia do apoio da Fundação para a Ciência e Tecnologia (Bolsa SFRH/BPD/35862/2007).

mano médio não terá mais do que 70% de hipóteses de fazer uma identificação correcta após 5 minutos de interrogatório.

Passados mais de cinquenta anos, o teste de Turing sofreu as mais diversas críticas: que a máquina não imita, estando a imitação apenas «no olho do observador»; que isso se deve à tendência dos humanos para atribuir estados mentais aos outros humanos e, por extensão, a comportamentos de outras entidades; que não basta observar o comportamento para fazer um juízo, sendo necessário conhecer os mecanismos de processamento interno; que o teste é apenas uma luta entre humanos (o programador da máquina e o interrogador); que só considera formas humanas de inteligência (quem só conhecesse gaivotas era capaz de pensar que só o seu padrão de voo é voar); que o interrogador pode sempre desmascarar a máquina se apelar à estrutura cognitiva de baixo nível, que depende de associações conceptuais que não se adquirem explicitamente (Saygin *et al.*, 2000).

Uma versão popular do teste de Turing são *robots de software* que «conversam» com humanos. Um precursor dessas experiências é o programa ELIZA, desenvolvido por Joseph Weizenbaum (1966): o utilizador humano escreve as suas «falas» no teclado e recebe do computador respostas também escritas com tempos de reacção que não desmentem a humanidade do interlocutor. Na versão mais conhecida a máquina desempenha o papel de um psicoterapeuta rogeriano, que devolverá sistematicamente as falas ao paciente com pedidos de esclarecimento. Weizenbaum explica que escolheu o psiquiatra a entrevistar por lhe parecer uma das poucas situações em que parece natural a pose de quase completa ignorância acerca do mundo real. Se o paciente diz «Fui dar uma grande volta de barco» e o psiquiatra responde «Fale-me de barcos», não pensamos que ele seja ignorante acerca de barcos, mas que ele está a perscrutar a nossa mente. São relatadas diversas histórias acerca da forma como muitas pessoas, interagindo com o programa, se convenciam de que consultavam um psicoterapeuta. Uma das secretárias de Weizenbaum terá chegado a pedir que a deixassem a sós com o «psicoterapeuta» para dispor da necessária privacidade. Contudo, como Weizenbaum explicou, o ELIZA não tinha qualquer forma de compreensão: assentava na detecção de palavras-chave armazenadas, em esquemas de decomposição das frases inseridas pelo humano e sua recomposição para produzir respostas, em regras simples para quebrar impasses. Weizenbaum é claro: quem atribui conhecimento e inteligência ao seu interlocutor é o humano; além dos truques relativamente simples da operação interna do ELIZA, tudo o resto é fornecido pelo humano utilizador.

Mais ambicioso foi o sistema apresentado por Terry Winograd: o SHRDLU deveria interagir com um humano na base da genuína compreensão da linguagem natural (Winograd, 1971). A ideia de partida é que, para isso, os computadores têm de saber do que estão a falar: conhecer e raciocinar com factos acerca do mundo. Winograd vai construir um mundo para um *robot* agir nele e sobre ele conversar. Existem neste mundo os seguintes objectos: o *robot* e um interlocutor; o *robot* tem uma mão e manipula objectos em cima de uma mesa, onde está uma caixa; os outros objectos são sólidos geométricos («blocos»): cubos, pirâmides, bolas. Porque o *robot* tem um olho, o sistema dispõe de informação completa acerca deste mundo. Podemos dar ordens em inglês (através do teclado) e o *robot* executará essas

ordens e fornecerá as informações que lhe solicitemos. Podemos ver a operação do *robot* num ecrã. Este é o «mundo dos blocos»: um programa num computador.

O que temos aqui? Qualquer pergunta que não seja sobre o micromundo dos blocos nunca desencadeará uma conversa. Mesmo sobre a sua «especialidade», o SHRDLU ignora quase tudo (nem sabe que os sólidos têm peso e ele pode variar com o tamanho). As relações básicas deste «mundo» são relações espaciais entre objectos; os únicos eventos possíveis são «acções» do *robot* manipulando os sólidos. Expressa-se assim a ideia de que podemos cortar fatias do mundo à nossa vontade e ele continuar a fazer sentido. O SHRDLU faz a economia da dinâmica da relação entre percepção e mundo: parece que vemos, nós e o «olho» do *robot*, certas coisas no ecrã do computador e que depois elas são descritas, mas é o contrário que acontece — o programa contém certas descrições (o cubo encontra-se numa certa posição) e o que aparece no ecrã é o resultado dessa descrição: «ser» é «ser descrito». Quando o utilizador dá uma instrução, o *robot* elabora o plano da sua execução e só começa a executar depois de concluir o plano e o verificar. Esse raciocinar é completamente separado do agir enquanto o plano não tiver sido adoptado — mas, uma vez adoptado o plano, ele coincide perfeitamente com a sua execução. Dá-se uma perfeita coalescência de mental e comportamental, de teoria e prática. Tudo o que se pode fazer pode antecipar-se (tem de antecipar-se). Nada do que não pode planificar-se à exaustão pode ser realizado. Isso só é possível porque «mundo» e «descrição do mundo» são o mesmo. Claro: o mundo não existe! Está tudo «dentro da cabeça» do *robot* (que, aliás, também não existe).

Winograd achava que as frases expressas em linguagem de programação ganhavam uma semântica genuína em linguagem-máquina por esta ligar directamente «ao mundo», isto é, aos estados físicos mais elementares do computador. Na verdade, interpretar desse modo a frase «Boise é uma cidade» é dizer que a expressão «BOISE» aponta para um endereço de memória com o rótulo «CITY». Isso não é mais do que pretender que «Napoleão venceu a batalha de Waterloo?» quer dizer «Verifique se a frase “Napoleão venceu a batalha de Waterloo” ocorre no volume que tem o número XXX,XXX na numeração decimal de Dewey na Secção da Rua 42 da Biblioteca da Cidade de Nova Iorque» (Fodor, 1978, pp. 204-211).

O «argumento do quarto chinês» (Searle, 1980) ajuda a ver o que está aqui em causa de forma geral. Suponhamos que temos um humano, falante de inglês e absolutamente ignorante de chinês, fechado num quarto. Passamos-lhe descrições de uma situação típica (uma ida ao restaurante), de uma ocorrência particular dessa situação (numa dada ida ao restaurante serviram-me um bife queimado, levantei-me e saí sem pagar) e perguntas acerca da história (comi o bife ou não?) — tudo escrito em chinês. Para o humano encerrado no quarto são apenas listas de caracteres incompreensíveis — mas reconhecíveis pela forma. Então passamos para dentro do quarto instruções, em inglês, acerca de como manipular as três listas de símbolos. Quando uma «pergunta» é passada ao humano dentro do quarto, as instruções em inglês dizem-lhe como organizar uma sequência de símbolos tirados das listas fornecidas para «responder» aos humanos cá fora. Com alguma habilidade na manipulação, essas «respostas» são indistinguíveis das de um falante de chinês. O

ponto é que o humano dentro do quarto é como um computador a funcionar com um programa: sabe seguir as instruções para manipular os símbolos, que distingue pela forma, mas nada sabe do seu significado. O problema é pretender-se que seja possível fixar uma imagem do mundo numa versão instantânea e estática da linguagem e pô-la a correr num mecanismo separado desse mesmo mundo, fechado num computador.

Parece, pois, que falta mundo às máquinas e que, sem algum mundo partilhado, será difícil conversar com elas. Mas a dificuldade da conversa é coisa que só nos acontece com as máquinas?

Podemos falar com um humano?

Durante a campanha para as eleições legislativas de 2005 foi publicado (*Público*, 22 de Janeiro de 2005, p. 2) o seguinte relato: «O líder social-democrata apelou ao eleitorado que votou PSD nas legislativas de 2002: “Ninguém que confiou em nós tem razão para deixar de confiar.” Mas esta frase foi toldada por um terrível *lapsus linguae*. Na realidade, o que se ouviu foi Santana dizer: “Ninguém que confiou em nós tem razão para deixar de desconfiar”.» O que é que permite à jornalista escrever que aquele dirigente partidário afirmou uma coisa apesar de o ter ouvido dizer o contrário?

A resposta está, claro, nos elementos contextuais que se misturam com o significado codificado linguisticamente. Contudo, o contexto não é um palco preparado com antecedência para receber os actores. Muitos factores podem mudar o contexto cognitivo. Sperber e Wilson (2001, pp. 189-259) propõem compreender essa dinâmica a partir da relevância. A relevância de uma elocução está ligada ao efeito que ela consegue obter em função das ligações contextuais e do processamento cognitivo requerido para activar essas ligações (um potencial elevado de efeito contextual pode nunca ser realizado por uma elocução cuja compreensão dependa de um pesado exercício dedutivo). A teoria da relevância assume a precedência da presunção de relevância em todo o acto de comunicação ostensiva. O falante, pelo próprio acto de falar, sugere que a sua fala é relevante; o ouvinte pressupõe que o falante pretende ser relevante. Todas essas presunções podem depois ser frustradas, mas não podemos eximir-nos a elas para começar quando comunicamos. Radicaria aí a nossa capacidade para identificar um *lapsus linguae* — e a nossa capacidade mais geral para falar com humanos sem desentendimentos generalizados.

E donde vem o diferencial de relevância? O sistema de referência da linguagem é a acção que os humanos têm em comum. As concordâncias requeridas por um determinado jogo de linguagem emergem da concordância de formas de vida (Wittgenstein, 1958, §240-241). Ora, a par da divisão social do trabalho, há também uma divisão do trabalho linguístico. «Ouro» refere-se a um elemento que quimicamente puro é quase branco, mas cujo estereótipo inclui «amarelo» porque o encontramos normalmente misturado com cobre. Nem todos os que sabemos o que é ouro o reconhecemos rigorosamente, mas podemos evitar comprar ouro falso recorrendo a especialistas cujos métodos pertencem, de forma colectiva, a todos — mas não a cada um — dos que adquiriram a palavra: «é apenas o estado sociolinguístico do corpo linguístico

colectivo a que o falante pertence que fixa a extensão» (Putnam, 1975, p. 229). Mais do que isso: a linguagem é produzida historicamente, em sociedade, na *praxis* de que ela é também ferramenta — e onde radica a relevância diferencial. Entretanto, a produção social da linguagem está vedada aos programas de computador, que estão radicalmente sós. E a nós, humanos, que laços nos mantêm ancorados historicamente a certas formas de vida? Essa pergunta leva-nos à questão da memória: à nossa e à dos computadores.

Quando pensamos na «memória» dos computadores somos em geral induzidos a conceber uma espécie de «armazém» onde estão guardados registos, que podemos aceder de modo a poder recuperar algum tipo de descrição objectiva do passado. Contudo, pelo menos do ponto de vista do acesso a esses registos, essa visão enfrenta dificuldades. Vejamos um caso prático. Como é que os utilizadores de computadores arquivam e recuperam ficheiros? Na maior parte dos casos orientam-se pelo paralelismo com a localização física e ignoram tanto quanto podem as ferramentas de «procura lógica» (procura por texto em nomes de ficheiros ou nos documentos). Por que é tão difícil que um utilizador reconheça a lógica que outro seguiu na organização da informação? Por que é que o próprio criador de uma determinada organização da informação tem, mais tarde, dificuldade em reconhecer espontaneamente as suas opções anteriores (e, portanto, dificuldade em encontrar os seus próprios ficheiros)? Chalmers (1999) sugere que só compreenderemos o que se passa se assumirmos que (também) as nossas práticas técnicas envolvendo objectos informáticos são mediadas pela linguagem. Uma palavra que usamos nos sistemas de gestão de informação, tendo-se tornado um símbolo, deixou de ser uma etiqueta ligada de forma lógica e absoluta a uma entidade determinada no mundo objectivo, que autorizaria uma reconstituição posterior da ligação fixa palavra/coisa. Trata-se de compreender que a relação entre um nome na linguagem e uma coisa no mundo tem um significado essencialmente diferencial: depende da estrutura das suas relações e diferenças com outros elementos da linguagem e da realidade, marcadas pela variação das circunstâncias, dos utilizadores, das actividades. É por isso que Chalmers, como alternativa ao paradigma da «recuperação», propõe, em parte inspirado no sistema de recomendações da livraria Amazon.com, a «filtragem colaborativa», na qual a navegação entre objectos num domínio é apoiada em indicações provenientes do mundo dos seus utilizadores, construídas com base na acumulação e cruzamento de indicações subjectivas múltiplas, corrigíveis, modificáveis com o tempo e as actividades correntes, decorrentes em parte do perfil dos objectos e em parte do perfil dos sujeitos. Isto é: importamos relações sociais para o mundo dos objectos (que as não têm) a partir do mundo dos sujeitos (que as têm) e assim podemos orientar-nos melhor.

A filtragem colaborativa, que consiste em nos apoiarmos nas relações sociais entre humanos para lidarmos melhor com objectos, tem uma curiosa contraparte na conceptualização da memória colectiva das sociedades humanas como uma dinâmica social suportada em objectos. É o caso de Kantschneider (2002), que concebe a memória colectiva como assente num triângulo hermenêutico entre artefactos, os seus produtores e os seus consumidores. Os artefactos importam porque a memória assente apenas nas comunicações inter-individuais é instável e precível. A interacção nua de indivíduo a indi-

víduo, sem objectivação (textos, rituais, imagens, edifícios, monumentos), não tem os meios para a memória colectiva. Do lado dos agentes, há que contar com os produtores dos artefactos, mas também, porque existe um problema de recepção, com os seus consumidores. Como «não há nenhuma ligação natural, directa, entre o real e o recordado», esses artefactos podem, enquanto representações, nunca passar do modo potencial ao modo real, constituindo-se em memórias falhadas. Por exemplo, nos EUA a guerra da Coreia é uma «guerra esquecida» entre as muito lembradas participações na Segunda Guerra Mundial e na guerra do Vietname.

De acordo com autores como Hirst e Manier (2008), compreender o que é a memória colectiva passa por compreender a sua inseparabilidade da construção da identidade colectiva. As «guerras da memória», exibindo em pleno factores sociais e políticos, deixam ver o que isso significa, como mostra o seguinte exemplo. Poucos anos depois da derrota dos EUA no Vietname, foi decidido nesse país construir um memorial aos respectivos veteranos. Aberto um concurso de projectos para o efeito, o Memorial foi construído segundo a ideia vencedora de Maya Lin: uma enorme estrutura de granito polido na forma de um V alongado, semi-enterrada, com a inscrição dos nomes das vítimas. Uma espécie de grande lápide tumular. Mas a polémica incendiou-se: alguns consideravam que o «V» parcialmente enterrado lembrava uma cicatriz e parecia recordar mais as fracturas políticas sobre a participação na guerra do que propriamente o esforço militar. Criaram-se organizações contra e a favor das diferentes visões do que devia ser o memorial. O resultado foi a construção de uma «adenda» ao memorial: uma estátua mais convencional representando três soldados de origens diferentes: um branco, um negro, um hispânico. Posteriormente outra questão emergiu: então e as mulheres que fizeram a guerra? Daí nasceu no princípio dos anos 1990 o terceiro componente: o Memorial às Mulheres na guerra do Vietname.

As guerras da memória, exibindo a relação entre memória colectiva e identidade colectiva, entre a forma como recordamos o passado e como tratamos do futuro, mostram que, apesar das tentações objectivistas, talvez não haja um passado comum imediatamente disponível. Assim, se parece difícil conversar com as máquinas por elas não terem um mundo, pode ser que, bem vistas as coisas, também seja difícil conversar com os humanos: a conversa pode ser ameaçada por *lapsus linguæ* insidiosos, formas de vida discordantes, disfunções da divisão social do trabalho linguístico, até pelo difícil acesso a um passado comum. Não obstante, em inúmeras ocasiões os desencontros radicais não perduram. Como pode isto ser? Como pode não haver conversa com uma máquina se nós a escutamos, mesmo que através dela fale outro humano agora ausente, ou fale o nosso passado? Podemos tomar essa máquina como um mero canal anódino? Se a memória colectiva é suportada em artefactos, e são esses artefactos inertes que dão tempo e estabilidade ao testemunho vivo do agente, o que vale a oposição subjacente? Como julgar morto o arquivo quando estamos sempre em risco de querer e não conseguir recuperar a sua máscara fúnebre, porque nos responde no tom do presente, no tom da nossa nova vida e não naquele que tínhamos quando arquivámos? Como julgar morta a letra escrita se ela, multiplicando os leitores e os palcos, se torna desse modo polissémica em mais alto grau? Como

encarar negativamente o facto de intervir algum procedimento mecânico na criação do arquivo, como se isso lhe sorvesse a vida, se basta colocar-nos na óptica de quem o redescobre para o entendermos como capaz de resistir aos nossos caminhos e nos levar por outros? Porquê tantas oposições em jogo? Provavelmente porque lemos o mundo como se ele fosse o que a consciência constituinte dele faz. Talvez dando prioridade ao mundo, e não à consciência, possamos ultrapassar estas oposições. E alguns dualismos. Tentemos esclarecer este ponto com a ajuda da noção de instituição, tomando-a de Merleau-Ponty.

Instituição.

Merleau-Ponty (1954) proporciona-nos um ponto de apoio para esclarecer esta dificuldade. O que está em causa filosoficamente nesse texto é a recusa de uma consciência constituinte, recusa da ideia de que só descobrimos no mundo aquilo que foi constituído pelas operações da nossa consciência: O que se opõe a «constituição» é «instituição». «Instituição significa estabelecimento numa experiência (ou num aparelho construído) de dimensões (sistema de referência) em relação às quais toda uma série de outras experiências terão sentido e formarão uma sequência, uma história.» (Merleau-Ponty, 1954, p. 38)

Merleau-Ponty fala-nos da instituição num plano mais vasto do que o habitual. Na ordem do vivo, na medida em que o desenvolvimento do organismo não resulta apenas do desdobramento de uma estrutura inata, mas também de uma plasticidade (entre os limites da espécie e da monstruosidade) que responde às contingências de tempo e lugar. Fala-nos da puberdade em termos de instituição da própria pessoa, lembrando que a instituição não liquida absolutamente o que a precedia. Fala-nos da instituição de um sentimento: um amor fica ligado aos acontecimentos contingentes onde começa a formar-se, cresce, transforma-se, morre. Fala-nos da instituição de uma obra, como no pintar ou no escrever: no princípio há um vazio; depois o projecto começa mas não está desde logo formado na cabeça do artista, vai sendo reconhecido nas realizações parciais; depois vai-se alimentando dos encontros com os outros. Por exemplo, um livro, que mostra como «toda a instituição tende para a série», deixa ver a instituição como instalação da diferença, desvio dentro da norma, fazer de uma nova norma possibilitadora de novos desvios. Fala-nos da instituição de um saber contra a ideia do saber como apropriação de ideias eternas, quando o saber é ele próprio histórico.

Enriquecida deste modo, a noção de instituição é então pensada na forma mais familiar: os corpos do Estado, as instituições sociais. Também aí, «a instituição não é apenas o que foi fixado por contrato, mas isso mais funcionamento» — tal como «a pintura não é a lógica da pintura», já que o pintor faz escolhas, mas não lhes faz a teoria; avança por meio de certos afastamentos expressivos em relação a uma certa norma, mas isso não é necessariamente logo uma tese. A instituição é universalizante, porque torna possíveis séries de acontecimentos com uma certa unidade, torna possível uma historicidade; mas também é particularizante, porque estabelece diferenças. Pela insti-

tuição, a história torna-se relação entre pessoas mediada por coisas. É essa materialidade das instituições é importante, porque as ideias dos humanos só existem nos aparelhos que as fixam e nenhum as fixa inteiramente. Pela instituição o sentido deixa de ser do domínio da minha interioridade: o sentido é depositado; o instituído tem sentido sem mim. A instituição instala-nos de uma certa maneira na história: nem estamos fechados no nosso presente (podemos penetrar no horizonte dos outros), nem nos elevamos nunca acima da história como espectador absoluto; não há criação pura e incondicionada (nisso erra o relativismo cultural radical), mas as instituições não são apenas limites, também abrem possibilidades.

O texto de Merleau-Ponty permite compreender a contribuição das máquinas que andam pelo nosso mundo para o instituir em curso, já que ela não depende da consciência que elas tenham disso. De modo mais geral, ajuda a articular uma agenda contra alguns dualismos: entre o interior e o exterior do indivíduo, entre o indivíduo e o colectivo, entre o corpo e a inteligência, entre o sujeito e o objecto, entre o presente e o imaginário, entre o natural e o cultural, entre o passivo e o activo, entre o mecânico e o criativo. Entre o dito e o por dizer. E talvez ajude a desmontar também algumas fronteiras entre escrita, memória e arquivo. Adivinha-se, contudo, a permanência de uma resistência central: nada disto pode resolver a insuficiência das máquinas para a semântica. Nesse ponto não fazemos apostas — fazemos experiências. Vejamos.

Robótica Institucionalista.

A Robótica Institucionalista, uma experiência sobre Sociedades Artificiais em curso no ISR/IST, inspirada na Economia Institucionalista e na noção merleau-pontiana de instituição, é uma estratégia para conceptualizar sistemas de múltiplos *robots* que toma as instituições como a ferramenta principal da vida social. A ideia de base é que para compreender uma sociedade não podemos focar apenas os indivíduos e as suas razões: temos de considerar o seu acoplamento com o respectivo ambiente físico e social, no qual se acumulam modificações introduzidas (intencionalmente ou não) por muitas gerações, modificações preservadas (ou simplesmente subsistentes) por passarem por ser possibilitadoras das formas de vida aí instaladas. Aí entra uma noção alargada de instituição. A Robótica Institucionalista é também uma forma de conceber ambientes partilhados por humanos e *robots* onde certas interacções sejam reguladas por instituições comuns. É por isto que os conceitos fundamentais não distinguem entre aplicações possíveis com *robots* ou com humanos: é o caso da definição de instituição que passamos a discutir.

Propusemos (Silva, Ventura e Lima 2008) a seguinte definição (provisória) de instituição: «As instituições são conjuntos cumulativos de modificações artificiais e persistentes do ambiente ou dos mecanismos internos de um subconjunto de agentes, tidas como funcionais para a ordem colectiva.» Esta definição pode ser instanciada, nomeadamente, por organizações ou grupos informais, convenções, normas, papéis, rotinas comportamentais, formas estereotipadas de perceber certas situações, artefactos materiais. Vamos procurar aclarar

essa proposta contrastando-a com um exemplar da classe de definições, aliás prevalentes, que consideram as instituições como sistemas de regras.

Hodgson (2006) define as instituições como sistemas duradouros de regras sociais estabelecidas e embutidas que estruturam as interações sociais. Esta definição opõe-se, por um lado, à pretensão, equivocada, de que as instituições (línguas, leis, moeda, sistemas de pesos e medidas, organizações) poderiam ser reduzidas a padrões de comportamentos correlacionados: é que a monarquia não deixa de existir quando todos os membros da família real estão a dormir. Por outro lado, contudo, Hodgson está ciente de que o papel das regras em contextos sociais não é redutível a mecanismos deliberativos: necessita englobar na explicação a base informal dos comportamentos estruturados e duradouros. Um elemento essencial desse alargamento é a noção de hábito: uma disposição para, na presença de um estímulo ou contexto adequado, iniciar comportamentos ou pensamentos anteriormente adoptados ou adquiridos. Hábitos persistentes, partilhados e prevalentes num grupo são a base dos costumes; hábitos enquadrados por ambientes organizacionais resultam em rotinas, algo como hábitos colectivos (Hodgson, 2007, p. 111). São os hábitos que permitem compreender a ligação entre o funcionamento das instituições e o comportamento dos indivíduos, graças à causalidade descendente reconstitutiva envolvida nas instituições: a capacidade das instituições para (de modo não determinista) mudar os hábitos de pensamento e de acção dos indivíduos. Mas hábitos e comportamentos são coisas diferentes: não estamos sempre a utilizar um hábito só porque o adquirimos. Vejamos algumas consequências desta visão.

Um aspecto crucial da proposta de Hodgson é que as instituições não são do domínio do natural: quando as leis da natureza forçam a agir de forma determinada não há lugar a normas ou sanções para produzir um determinado comportamento. O mecanismo de replicação das instituições é ele próprio social, excluindo a transmissão genética: a existência (potencial ou efectiva) de sistemas de regras muito diferentes em ambientes naturais similares é um indicador de que estamos a lidar com mecanismos de transmissão socioculturais, não genéticos. Contudo, as leis da natureza restringem o domínio institucional (uma regra que exige do agente que ele desafie as leis da gravidade está fora desse domínio). Ora, ao reconhecer que o conjunto das regras possíveis pode ser alargado, por exemplo por desenvolvimentos tecnológicos que tornem possíveis acções que antes não o eram, parece que Hodgson tem de admitir que a relação entre natureza e instituições não é estável: a fronteira existe mas é móvel. A definição de instituição devia, a nosso ver, acomodar essa compreensão.

O que nos pareceria adequado, neste caso, seria que a definição de instituição não tentasse traçar uma fronteira demasiado rígida entre mecanismos externos (regras e dispositivos associados) e mecanismos internos dos agentes (por exemplo, hábitos). Até porque, uma vez que as representações mentais de uma instituição são em parte constitutivas dessa instituição (Hodgson, 2006, p. 4) — por exemplo, devo reconhecer um contexto para poder agir da forma apropriada a esse contexto —, certos mecanismos internos dos agentes (mentais e sensoriais) são inseparáveis do mecanismo institucional. E, como nem todos são determinados geneticamente (a aprendizagem, por

exemplo), podemos modificar parcial e limitadamente os mecanismos internos dos agentes e, por essa via, o funcionamento institucional. Para falar nos termos das ciências do artificial, podemos dizer que isso se obtém modificando os modelos do mundo dos agentes (ou o seu aparato sensorial). Mas também podemos pensar nas ideologias dos humanos como variações coletivas (em subconjuntos de indivíduos de uma mesma população) dos modelos do mundo em uso numa dada sociedade.

Ora, é a essa exigência que responde a definição de instituição que propomos — e por isso a entendemos preferível às que concebem as instituições fundamentalmente como sistemas de regras sociais.

Conclusão em aberto.

As nossas reflexões acerca da linguagem e da memória, no quadro da investigação sobre Sociedades Artificiais, conduziram-nos a uma valorização dos artefactos que povoam o mundo em instituição. *Robots* e humanos, tão diferentes na óptica da consciência constituinte, talvez sejam afinal parceiros no instituir. Na medida em que nos tornemos herdeiros das modificações que introduzem no mundo em que nos cruzamos. A experiência continua. «Fazer filosofia com uma chave de fendas», como disse um dia Inman Harvey.

Chalmers, Matthew

1999 «Structuralist Informatics: Challenging Positivism in Information Systems», in *Proceedings of the Conference of the UK Academy of Information Science*, Londres, McGraw-Hill, 1999, pp. 13-22.

Fodor, Jerry A.

1978 «Tom Swift and his Procedural Grandmother», conforme republicação in Fodor, Jerry A., *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Brighton (Sussex), The Harvester Press, 1981, pp. 204-224.

Hirst, William e Manier, David

2008 «Towards a Psychology of Collective Memory», *Memory*, vol. 16, n.º 3, pp. 183-200.

Hodgson, Geoffrey M.

2006 «What Are Institutions?», *Journal of Economic Issues*, vol. XL, n.º 1, Março de 2006, pp. 1-25.

2007 «Institutions and Individuals: Interaction and Evolution?», *Organization Studies*, vol. 28, n.º 1, pp. 95-116.

Kansteiner, Wulf

2002 «Finding Meaning in Memory: A Methodological Critique of Collective Memory Studies», *History and Theory*, n.º 41, pp. 179-197.

Merleau-Ponty, Maurice

1954 «L'institution dans l'histoire personnelle et publique», in Merleau-Ponty, Maurice, *L'Institution: La Passivité. Notes de cours au Collège de France (1954-1955)*, s. l., Belin, 2003, pp. 31-154.

Putnam, Hilary

1975 «The Meaning of "Meaning"», in *Mind, Language and Reality (Philosophical Papers, volume 2)*, Cambridge, Cambridge University Press, 1980, pp. 215-271.

Saygin, Ayse Pinar, Cicekli, Ilyas e Akman, Varol

2000 «Turing Test: 50 Years Later», *Minds and Machines*, vol. 10, n.º 4, pp. 463-518.

- Searle, John R.
1980 «Minds, Brains and Programs», *Behavioral and Brain Sciences*, n.º 3, pp. 417-424.
- Silva, Porfírio, Ventura, Rodrigo e Lima, Pedro U.
2008 «Institutional Environments», in Jung, B., Michel, F., Ricci, A., e Petta, P. (orgs.), *From Agent Theory to Agent Implementation: 6th International Workshop, Working Notes*, 13 de Maio de 2008, AAMAS 2008, Estoril, pp. 157-164.
- Sperber, Dan e Wilson, Deirdre
2001 *Relevância: Comunicação e Cognição*, Lisboa, Fundação Calouste Gulbenkian.
- Turing, Alan M.
1950 «Computing Machinery and Intelligence», *Mind*, n.º 59, pp 433-460.
- Weizenbaum, Joseph
1966 «ELIZA: A Computer Program for the Study of Natural Language Communication between Man and Machine», *Communications of the ACM*, vol. 9, n.º 1, pp. 36-45.
- Winograd, Terry
1971 *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language (Artificial Intelligence Technical Report n.º 231)*, Cambridge (MA), MIT.
- Wittgenstein, Ludwig
1958 *Investigações Filosóficas*, in *Tratado Lógico-Filosófico, seguido de Investigações Filosóficas*, Lisboa, Fundação Calouste Gulbenkian.