

DIOGO FERNANDES*

CONVENÇÕES, CONHECIMENTO COMUM E CONVENÇÕES ALTERNATIVAS

Resumo

O objectivo deste artigo consiste em averiguar se a condição do conhecimento comum que, segundo David Lewis, define uma convenção é realmente necessária para explicar todos os mecanismos envolvidos no fenómeno social assim habitualmente designado, e se a definição de Lewis vai ao encontro das nossas intuições fundamentais acerca do significado do conceito. Para esse efeito, a mesma será testada na tentativa de encontrar resposta a duas questões: Qual a origem das convenções? Como se mantêm as convenções operacionais ao longo do tempo? Em 1. apresentarei os conceitos de «conhecimento comum» e de «saliência», e tentarei mostrar como ambos se encontram estreitamente relacionados de modo que um não pode ser concebido sem o outro aquando da explicação da origem das convenções. Em 2. apresentarei a crítica de Ken Binmore à possibilidade de um evento se tornar conhecimento comum numa população alargada, inserindo essa crítica no contexto da literatura da teoria dos jogos, elaborada posteriormente sobre a definição original de Lewis. Em 3. apresentarei alguns exemplos de Tyler Burge que questionam a necessidade de incluir na definição de convenção a cláusula do conhecimento comum de convenções alternativas. Em 4. farei um balanço das críticas apresentadas e concluo que a condição do conhecimento comum, enfraquecida mas preservada no essencial, é necessária para uma explicação satisfatória dos mecanismos em causa, permitindo uma resposta cabal às duas questões mencionadas.

Palavras-chave: convenção, saliência, conhecimento comum, truísmo, função social

Abstract

In this paper I'll try to answer the question about whether the common knowledge condition, which according to David Lewis defines a convention, is really necessary to explain the social phenomenon thus designated, and also whether Lewis' definition accords with our common intuitions about this notion. In order to achieve my goal I'll test the definition by asking two questions: Can the definition explain the origin of conventions? And can it explain the way conventions maintain themselves in use over time? In 1. I'll present the notions of common knowledge and salience, and I'll try to show how both are strictly connected in a way that one cannot be conceived without the other. In 2., following

* Bolseiro de Doutoramento da FCT. Licenciado em Filosofia pela Universidade de Lisboa, Pós-Graduado em Filosofia Analítica pela Universidade de Lisboa e Mestre em Filosofia Moderna e Contemporânea pela Universidade do Minho. Centro de Filosofia da Universidade de Lisboa – LanCog..

the literature on game theory developed after and over Lewis' definition, I'll introduce Ken Binmore's criticisms about whether it is possible for an event to become common knowledge in a large population. In 3. I'll introduce some examples by Tyler Burge which question the need to include, in the definition of convention, the proviso according to which there has to be common knowledge of alternative conventions. In 4. I'll evaluate the arguments presented and conclude that a weakened version of the proviso is necessary to give a complete answer to the relevant questions.

Key-words: convention, salience, common knowledge, truism, social function

1. Conhecimento comum e saliência

David Lewis¹ mostrou como é possível estabelecerem-se convenções sociais sem que seja necessário existir um acordo explícito entre os indivíduos de uma população que sejam parte dessa convenção. Tal não elimina, contudo, a possibilidade de o acordo explícito ser uma das maneiras mais eficazes de estabelecer uma convenção.

Quando no final de um jantar de ex-colegas de liceu, um deles propõe que se passem a reunir naquele mesmo restaurante de seis em seis meses, e cada um dos participantes dá o seu assentimento, estabeleceu-se por essa via uma convenção: uma regularidade comportamental R, entre os membros de uma população P, tal que

- 1) cada um dos membros de P passa a fazer seja qual for o conteúdo de R (consistindo R numa descrição da acção que cada um dos membros de P executa enquanto parte dessa convenção);
- 2) cada um dos membros de P tem a expectativa que cada um dos outros membros de P execute R;
- 3) cada um dos membros de P prefere fazer R na condição de cada um dos outros membros de P também o fazer.

A grande originalidade do tratamento de Lewis consistiu em mostrar detalhadamente quais são os mecanismos que permitem que se estabeleçam estas regularidades comportamentais, no interesse de todos os envolvidos, quando não existe a possibilidade de comunicação directa. Ou seja, como se podem resolver problemas de coordenação através da seleção de equilíbrios que se destacam pela sua saliência. Contudo, tanto acerca dos casos de acordo explícito, como acerca destes últimos, coloca-se uma questão essencial: de que modo é possível estabelecer-se o sistema de expectativas mútuas concordantes descrito em 2, e quais as suas características, tal que este permita a todas as partes desenvolver um raciocínio que estabeleça, para cada uma delas, a conclusão de que possui uma razão decisiva para fazer R?

Como para cada um dos membros de P a sua preferência em fazer R é condicionalmente dependente de uma igual preferência por parte de cada um dos outros (como descrito em 3), é necessário cada um assegurar-se que cada um dos outros possui, de facto, essa preferência. Se cada um tiver uma razão para acreditar que todos os outros têm essa preferência e acreditar que, tal como ele próprio, também os outros possuem as suas capacidades elementares de

¹ D. LEWIS, *Convention. A Philosophical Study*, Cambridge, Harvard University Press, Massachusetts 1969, Caps. 1-3. Para a definição de Lewis e várias outras, ver também M. RESCOLA, «Convention», in *Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, URL = <http://plato.stanford.edu/entries/convention/>.

raciocínio, então, de um modo informal, o tipo de raciocínio que cada um desenvolve consiste em colocar-se no papel do outro, replicando, portanto, o seu próprio raciocínio. Ou seja, enquanto parte da convenção, eu ajo considerando as expectativas que possuo relativamente às expectativas que o outro possui acerca das minhas próprias expectativas, e assim sucessivamente. Significa isto que eu possuo expectativas acerca de expectativas, ou seja, expectativas de ordem- n em que $n \geq 1$.

A questão pode ser colocada nos seguintes termos: Que premissas justificam a conclusão de que existe em P um sistema de expectativas mútuas concordante? Para Lewis, essas premissas têm de ser eficazes para estabelecer que algo é conhecimento comum numa população P . Mas em que consiste exatamente algo ser conhecimento comum em P , digamos uma proposição A ? Por exemplo, cada um dos alunos de uma turma pode saber que «o Professor X é muito exigente», sem que isso constitua conhecimento comum. Pode acontecer, por exemplo, que cada um dos alunos pense que apenas ele sabe que «o Professor X é muito exigente». Assim, a proposição A é conhecimento comum em P se, e somente se, cada membro de P sabe que A é o caso, cada membro de P sabe que cada membro de P sabe que A é o caso, cada membro de P sabe que cada membro de P sabe que cada membro de P sabe que A é o caso, e assim *ad infinitum*.

Poderá parecer estranho que num dos lados desta bicondicional surja uma proposição cujas condições de verdade parecem impossíveis de estabelecer, nomeadamente um conjunto infinito de crenças ou, de outro modo, uma sucessão infinita de crenças de ordem- n . Contudo, a literatura em teoria dos jogos acerca do conceito de conhecimento comum está de acordo em que a maneira mais comum de definir um operador de conhecimento comum CK (o qual partilha as propriedades de um operador de conhecimento individual) é a seguinte: $(CK)E$ consiste num evento em que (cada um conhece)² E é verdadeiro para todos os valores de n .

Com esta definição em mãos torna-se agora possível analisar de forma mais adequada as condições que Lewis oferece para que algo se possa tornar conhecimento comum e relacioná-las com a sua teoria da saliência. Portanto, _____ é conhecimento comum numa população P se, e somente se, existir um estado-de-coisas A tal que:

- 1) Cada um dos membros de P tem uma razão para acreditar que A é o caso.
- 2) A indica a cada um dos membros de P que cada um em P tem uma razão para acreditar que A é o caso.
- 3) A indica a cada membro de P que _____³.

Uma interpretação destas condições passa pelo esclarecimento daquilo em que consiste A , _____, e a relação de indicar. No exemplo acima considerado, A consiste no estado-de-coisas, ou evento, em que todos os ex-colegas, na presença uns dos outros, dão o seu assentimento para com a proposta de um deles em se reunirem com certa regularidade, num certo local. Assim, A é aquilo que define o conteúdo de _____, em que _____ pode consistir numa descrição da regularidade comportamental que, a partir desse evento inicial, passa a constituir propriamente a convenção. Por seu lado, a relação de indicar deixa-se explicitar pelo conceito de saliência.

O conceito de saliência, no tratamento de Lewis, tem como papel responder à questão:

² Ver a definição em K. BINMORE, «Do Conventions Need to Be Common Knowledge?», *Topoi*, 27 (2008) 17-27.

³ D. LEWIS, op. cit., p. 56.

Qual a origem das convenções? O conceito tem origem num outro, de Thomas Schelling, denominado por «ponto focal»⁴. Para o ilustramos, consideremos um exemplo do próprio Schelling, que consistiu numa experiência real. Foi pedido a um conjunto de sujeitos que, sem comunicação prévia, tentassem encontrar-se num determinado dia em Nova Iorque, sem indicação do lugar ou da hora. Surpreendentemente, ou não, a maioria dos sujeitos conseguiu resolver com sucesso este problema de interação estratégica – um problema de coordenação, na linguagem de Lewis. A maioria resolveu estar ao meio-dia na Grand Central Station. Assim, a Grand Central Station foi o ponto focal, ou a saliência, que determinou o comportamento dos sujeitos e que constitui a base ou o elemento de conhecimento comum que lhes permitiu racionar através de replicações, i.e, com base num sistema de expectativas mútuas. É natural que, caso a experiência fosse conduzida hoje, o ponto focal viesse a ser diferente – talvez o «ground zero» do 9/11 ou Times Square. Se a experiência fosse realizada em Los Angeles, é provável que os sujeitos escolhessem a entrada dos estúdios da Paramount em Melrose Avenue. Mas, em 1960, numa época em que o comboio era o meio de transporte por excelência para viajar de, e para, Nova Iorque, um ex-libris da cidade que figurava com frequência nos cartazes turísticos da cidade numa época de ouro do *design* publicitário, a escolha dos sujeitos acaba por não ser surpreendente.

Como diz Schelling, a característica fundamental deste tipo de soluções para problemas de coordenação consiste na sua «conspicuidade», a qual pode assumir dimensões diversas que, de algum modo, apelam aos traços psicológicos subjectivos dos agentes envolvidos. Pode tratar-se, assim, de uma conspicuidade baseada em precedentes, em analogias, em propriedades intrínsecas, geométricas ou estéticas, da situação em causa, naquilo que os agentes sabem acerca uns dos outros, etc. Por exemplo, numa situação em que existem apenas dois agentes, se cada um deles souber que o outro é versado no estudo da teoria dos jogos, e cada um deles conhecer, e souber que o outro conhece, os «payoffs» de cada entrada da matriz que representa o seu problema de coordenação (ver os exemplos de matrizes na página seguinte), e dar-se o caso de essa matriz ter nove entradas (três linhas que correspondem às alternativas do jogador 1 e três colunas que correspondem às alternativas do jogador 2) em que os pontos de equilíbrio se encontram numa linha diagonal, cada um deles poderá optar pelo equilíbrio que corresponde à entrada central da matriz (R2, C2), por uma questão de saliência geométrica ou estética. A relação de indicar consiste, assim, no modo como uma solução possível para um problema de coordenação se torna conspícua para todos os sujeitos envolvidos nessa interação, i.e., no modo como algo de comum se torna para eles saliente.

Voltando às três condições de Lewis, *A* corresponde assim ao elemento saliente da situação - que definimos de modo a incluir o acordo explícito, como sendo o tipo mais conspícuo de saliência -, o qual, em virtude de o ser, «indica» a todos os envolvidos os factos relevantes para a resolução do problema de coordenação. De outro modo, *A* indica a todos os envolvidos os factos relevantes em virtude de *A* apelar para as características psicológicas de cada um que, por seu lado, «selecionam» *A* devido ao seu tipo específico de conspicuidade. Importa notar que, para estarmos na presença de um genuíno problema de coordenação, i.e., uma situação de interação estratégica em que a saliência desempenha um papel relevante na escolha de um equilíbrio, uma condição importante terá de se verificar: a estrutura racional da situação estratégica não pode ser suficiente para se alcançar o equilíbrio de coordenação. Isto pode suceder de duas maneiras, as quais Lewis pretende eliminar, de modo a definir um problema

⁴ T. C. SCHELLING, *The Strategy of Conflict*, Oxford University Press, New York 1963.

de coordenação. Primeiro, não pode existir uma estratégia dominante num problema de coordenação, caso em que a racionalidade dos agentes é suficiente, por si só, para se alcançar a coordenação. Considere-se, por exemplo, as seguintes matrizes⁵:

| | | | | | |
|----|-------|-------|----|-------|-------|
| | C1 | C2 | | C1 | C2 |
| R1 | (1,1) | (1,1) | R1 | (1,1) | (0,0) |
| R2 | (0,0) | (0,0) | R2 | (0,0) | (9,9) |

Dado que estamos perante jogos de informação perfeita, em que ambos os jogadores conhecem os «payoffs» e sabem que os outros também os conhecem – e, no caso de se tratar de um jogo sequencial, ambos têm acesso a todas as jogadas prévias – no primeiro deles o «row-chooser» tem R1 como escolha dominante: faça o que fizer o «column-chooser», R1 é sempre a melhor opção; logo, não importa a escolha que este último fizer, pois, dada a racionalidade do primeiro, um dos dois equilíbrios possíveis será sempre alcançado. No segundo jogo, dadas as condições da informação perfeita e da racionalidade elementar comum aos agentes – todos preferem o equilíbrio que maximiza os seus «payoffs» – o único equilíbrio próprio existente (R2, R2) será sempre escolhido. Em suma, um problema de coordenação verifica-se se, e somente se, existirem pelo menos dois equilíbrios próprios e, mesmo que tal seja o caso, não existam estratégias estritamente dominadas.

Falta referir um elemento essencial do raciocínio que conduz à formação de um sistema de expectativas mútuas concordantes. Apesar de termos indicado as premissas que permitem a cada membro de P acreditar que cada um dos outros tem uma razão para acreditar que *A* é saliente para cada um deles; e, como tal, uma razão para acreditar que se encontra estabelecida uma convenção; ainda não foram indicadas as premissas que permitem que cada membro de P acredite que cada uma dos outros acredita realmente que se encontra estabelecida uma convenção. Ou seja, se eu acreditar que os outros agentes envolvidos no problema de coordenação não racionam como eu, i.e., de forma a replicar o meu próprio raciocínio e o de cada um dos outros, então, como a minha preferência em fazer R é condicional, eu não terei uma razão para agir de acordo com R. Em suma, é necessário que eu acredite que cada um dos outros possui a mesma capacidade de raciocínio que eu, nomeadamente a mesma capacidade de raciocinar indutivamente e a mesma informação que eu tenho a respeito dos factos relevantes – conhecimento dos «payoffs», a mesma percepção relativamente à saliência, etc. Em suma, para cada expectativa ou crença de ordem-*n* deve existir uma crença ou expectativa auxiliar acerca da racionalidade dos outros membros de P de ordem correspondente. Assim, quanto mais se sobe na ordem das expectativas, mais exigentes terão de ser os graus de racionalidade que as premissas auxiliares atribuem aos outros membros de P. Daí que seja perfeitamente razoável acreditar que, na maior parte dos casos, os sistemas atuais de expectativas mútuas concordantes não alcancem ordens particularmente elevadas. Temos, assim, como base do conhecimento comum, todas as premissas que permitem gerar um sistema de expectativas mútuas concordantes: 1) aquelas que dizem respeito à saliência de um equilíbrio e ao modo como tal saliência indica aos membros de P razões para esperar e acreditar; e 2) as premissas que permitem aos membros de P atribuir uns aos outros graus de racionalidade de ordem-*n*.

Por último, também me parece razoável concluir que a força do primeiro tipo de premissas

⁵ D. LEWIS, op. cit., p. 21.

pode influenciar a força das do segundo tipo. Por exemplo, uma saliência que consista num único precedente será mais fraca do que uma saliência que consista nesse mesmo precedente e na sua repetição, e assim sucessivamente. Também a conformação para com uma convenção estabelecida há séculos dará origem a premissas ou expectativas mútuas mais fortes do que uma única repetição de um precedente. Logo, para efeitos de coordenação, uma saliência fraca parece exigir mais das capacidades de raciocínio dos membros de P - replicação e detecção de saliências - do que uma saliência particularmente forte, tal como a conformidade para com uma convenção centenária.

2. Conhecimento comum numa população alargada?

A definição Lewisiana de convenção pretende acomodar duas das nossas intuições fundamentais acerca do que são as convenções. Uma delas é que, devido à natureza dos problemas que tentam solucionar – problemas de coordenação com pelo menos dois equilíbrios próprios - as convenções sociais têm sempre, pelo menos, uma alternativa incompatível, ou seja, uma alternativa tal que seja impossível fazer R e R' simultaneamente, em que R' é a regularidade comportamental alternativa a R. A outra intuição comum acerca das convenções pode ser expressa de modo informal da seguinte maneira: todas as partes envolvidas têm de conhecer certos factos acerca umas das outras; por outras palavras, para que algo seja uma convenção, todas as partes têm de partilhar um certo conhecimento comum. Normalmente, a explicação mais satisfatória que encontramos para o falhanço de algumas interacções sociais – senão da maioria – consiste no facto de algumas das partes desconhecerem algo que as outras conhecem.

O desconhecimento das convenções pode ser interpretado de duas maneiras. Em sentido literal, as partes podem desconhecer as convenções relevantes. Assim, alguém que no Japão não saiba que a vénia é um elemento fundamental de cortesia, e que, como tal, não corresponda às vénias que lhe são feitas, ficará certamente bastante mal vista. Contudo, existe um outro sentido em que se pode dizer de alguém que desconhece uma convenção, um sentido mais técnico que apela para a definição de convenção dada por Lewis, segundo o qual alguém pode realmente exibir a regularidade comportamental exigida, embora o faça pelas razões erradas; ou, se quisermos, alguém que faça R ou R', mas que desconheça todas as cláusulas da definição de R. Uma tal situação parece ir contra a referida intuição. Por exemplo, alguém pode conduzir pelo lado direito da estrada pensando que é o único que o faz com consciência de que se encontra a cumprir uma condição necessária da segurança rodoviária - para a qual existe uma alternativa contrária verificada noutros países - e que todos os outros apenas o fazem por uma questão de hábito. E poderá ainda acontecer que a maioria, ou até mesmo todos os condutores pertencentes a uma dada população partilhem dessa crença errada e, no entanto, se verifique uma regularidade comportamental que em nada se distingue de uma idêntica regularidade em que todas as partes sabem que estão a cumprir a sua parte numa convenção.

Este exemplo de Lewis pode não parece muito convincente devido à implausibilidade de uma tal situação se verificar, ainda que me pareça realmente contraintuitivo falar da existência de uma convenção em tal cenário. Contudo, essa implausibilidade só vem fortalecer a posição de Lewis, pois quanto mais difícil for conceber a existência actual de uma regularidade comportamental que não seja baseada num sistema de expectativas mútuas, mais nos convencemos de que é necessário existir conhecimento comum numa convenção. A isto acresce o facto de que esse conhecimento, tal como o conhecimento individual que cada um

possui acerca de qualquer proposição, pode ser apenas do tipo potencial. Podemos ter ao nosso dispor a evidência necessária para chegarmos à conclusão de que as nossas regularidades comportamentais preenchem todos as condições que definem tecnicamente uma convenção, embora não tenhamos efectuado o raciocínio necessário para a alcançarmos. Além disso, este conhecimento, potencial ou não, pode ser do tipo não-verbal, o que sobressai nos casos em que os detalhes da regularidade em causa são de tal modo intrincados que só um especialista os conseguirá enunciar. Numa analogia com regras do futebol, muitos poderão não possuir as competências linguísticas ou analíticas necessárias para definir um «fora-de-jogo», embora, e correndo o risco de estar a ser demasiado optimista, a esmagadora maioria das pessoas saberá identificar um. Pode ainda ser possível não se ter uma crença geral acerca do que caracteriza uma situação S a que se aplica uma determinada convenção e, no entanto, saber identificar uma situação particular, quando e onde esta surge, e agir de forma adequada. Ou seja, é possível ser-se desprovido de uma crença do tipo «Espero que cada membro de P envolvido numa instância particular de uma situação S faça R» – que inclui uma quantificação universal sobre instâncias de S e membros de P – e, no entanto, saber identificar, por experiência adquirida, cada instância particular e desenvolver os comportamentos adequados. Ou seja, dada uma instância específica – conduzir de casa ao supermercado no dia tal-e-tal – eu sei que cada um dos outros condutores irão conduzir pela direita e que, como tal, prefiro conduzir também pela direita.

Todas estas qualificações acerca do tipo de conhecimento comum exigido para o estabelecimento de uma convenção permitem defender a posição de Lewis de um determinado argumento contra a necessidade de incluir esta exigência na definição. Se alguém propusesse a realização de um teste aos membros de uma população, em que se lhes pediria para identificar uma convenção e estes não o conseguissem fazer, ou negassem inclusive serem partes envolvidas, tal não seria suficiente para desclassificar uma regularidade comportamental como convenção, nem para se concluir que existem convenções que não implicam conhecimento comum.

Como já foi referido na secção anterior, Lewis oferece uma definição infinitista – se quisermos – de conhecimento comum, baseada numa hierarquia de expectativas e crenças de ordem-n. Uma questão que se coloca é a de saber se essa hierarquia infinita pode ser realmente gerada por um conjunto finito de premissas ou axiomas, os quais nos são dados pela definição de trabalho de Lewis⁶. Contudo, é necessário dar conta de outro problema que a definição enfrenta, o de saber se, de um ponto de vista epistémico, é realmente possível uma dada população P alcançar conhecimento comum a respeito de um evento ou uma proposição. Se se chegar à conclusão de que tal não é possível, pelo menos em alguns casos, então deveremos também concluir que existem convenções que não são conhecimento comum.

Com efeito, somos obrigados a admitir que nenhuma criatura finita pode conduzir um raciocínio que se estenda infinitamente ou que tal criatura mantenha um número infinito de estados epistémicos. O próprio Lewis o reconhece e contorna o problema afirmando que não nos encontramos realmente perante um raciocínio efectuado por indivíduos actuais, mas sim

⁶ R. AUMANN, em «Agreeing to Disagree», *Annals of Statistics*, 4 (1976) 1236-1239, apresentou um método finitista, baseado na teoria dos conjuntos, o qual supostamente permite gerar a hierarquia de conhecimento comum. Os autores dividem-se quanto à compatibilidade deste método com a caracterização informal de Lewis. Um dos autores já mencionados, Peter Vanderschraaft, por exemplo, argumenta no sentido de mostrar que tal é o caso, cfr. P. VANDERSCHRAAFT, «Knowledge, Equilibrium and Convention», *Erkenntnis*, 49 (1998) 337-369.

perante um conjunto infinito de implicações. O mesmo efeito é alcançado pela distinção entre razões para acreditar e crenças actuais, as primeiras consistindo propriamente no conteúdo da hierarquia infinita de implicações, e as segundas nas três, quatro, cinco ou seis replicações que o raciocínio dos seres humanos comuns consegue realmente alcançar. Não parece, assim, colocar-se um problema sério ao considerar-se a existência de uma hierarquia infinita de crenças potenciais, implicada pela hierarquia infinita de razões para acreditar. Note-se que «potencial» não tem aqui o mesmo sentido em que o termo foi utilizado para referir um tipo possível de conhecimento comum. «Potencial» poderia aí ser interpretado, senão como algo que está implícito, pelo menos como estando presente em potência, podendo tornar-se actual após devida reflexão. Aqui «potencial» qualifica algo que, pelo contrário, nunca estará ao alcance de uma criatura com poderes finitos de raciocínio.

Contudo, este não é o único problema que se prende com a possibilidade de acesso epistémico a este tipo de conhecimento. Uma crítica direccionada contra a possibilidade actual de algo vir a tornar-se conhecimento comum, ainda que este se encontre reduzido a uma hierarquia finita, foi avançado por Ken Binmore⁷. Para o compreendermos necessitamos de acrescentar um outro conceito àqueles que temos vindo a definir. De acordo com a literatura do conhecimento comum, um evento é conhecimento comum se, e somente se, for implicado por um truísmo, em que por «truísmo» se entende um evento que não pode ocorrer sem que cada membro de P saiba que esse mesmo evento ocorreu. Ou seja, um evento E é um truísmo – ou um evento público em P, se quisermos – se, e somente se, não poder ocorrer sem que $E = (\text{cada um conhece}) E$ ⁸.

Considere-se o exemplo dos alunos de uma turma em que cada um deles sabe que A, embora nenhum saiba que cada um dos outros sabe que A (ou, então, cada um sabe que todos sabem que A, embora nenhum saiba que todos sabem que todos sabem que A). Vimos que tal não configurava um caso de conhecimento comum. Alguma literatura designa este caso como sendo de conhecimento mútuo e atribui-lhe um estatuto de melhor aproximação possível à situação ideal, para efeitos de coordenação, de conhecimento comum. A questão a colocar é a seguinte: que tipo de evento público poderia fazer com que A se tornasse conhecimento comum nessa população específica? Se um dos alunos enunciasse, na presença de cada um dos outros, a frase «Disseram-me que o Professor X é muito exigente», então é razoável supor que tal evento – a enunciação verbal da frase – preencha os requisitos necessários. Considere-se uma outra situação bastante comum: um empregado de restaurante tropeça e deixa cair todo o conteúdo de um terrina de sopa no vestido de noite de uma cliente. De seguida, o empregado pronuncia imediatamente a frase: «A culpa foi toda minha, peço desculpa!» Com que intuito foi pronunciada essa frase? Era conhecimento mútuo entre o empregado e a cliente que a culpa tinha sido dele. A frase teve, pois, como objectivo fazer com que a cliente soubesse que ele desejava que ela soubesse que a culpa tinha sido dele. O resultado foi uma crença de 3ª ordem: a cliente sabe que ele sabe que ela sabe que a culpa foi dele⁹. Em suma, um certo estado mental – o desejo do empregado de tornar conhecida a sua crença da sua

⁷ Ver K. BINMORE, op. cit.

⁸ Para uma introdução exhaustiva ao conceito, ver P. VANDERSCHRAAF – G. SILLARI, «Common Knowledge», in *Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, URL = <http://plato.stanford.edu/entries/common-knowledge/>.

⁹ Parece-me que para se verificar *conhecimento comum* é necessário existirem, pelo menos, crenças de 3ª ordem. Mais, existindo crenças de 2ª ordem, e existindo a esse nível atribuição mútua de racionalidade,

própria culpa – tornou-se conhecimento comum numa população composta, pelo menos, por dois indivíduos.

Ora, a crítica de Binmore faz-nos notar o seguinte: se numa população de dimensão restrita é perfeitamente razoável pensar que um evento E se pode tornar conhecimento comum, quanto maior for a dimensão de uma população mais difícil será conceber a possibilidade de existir um evento tal que cada um dos membros de P observe cada um dos outros a observar E. Se pensarmos em convenções respeitadas por uma população de milhões, tal situação parece tornar-se impossível. Face a isto, parece necessário concluir que muitas das nossas mais banais convenções, seguidas por milhões de indivíduos, tais como, por exemplo, as convenções linguísticas, ou outras tão simples como conduzir pela direita, podem manter-se operacionais sem que se verifique a existência de conhecimento comum.

A definição final de Lewis encontra-se, todavia, enfraquecida num aspecto que parece, à primeira vista, facilitar uma resposta à crítica de Binmore. Existem graus de convenções que se distinguem pelas diferentes fracções da população que satisfazem as cláusulas da definição. Uma convenção, ou uma qualquer regularidade comportamental, passa a ser definida como mais ou menos convencional do que uma outra regularidade, consoante maior ou menor for a fracção populacional que satisfaz as cláusulas. Passamos a ter já não uma conformidade perfeita de todos os membros de P para com R, já não um sistema de expectativas mútuas que abrange todos os membros de P – mas apenas quase todos -, e já não um sistema de preferências condicionais que abrange todos os membros de P – mas apenas quase todos. Se esta definição parece corresponder àquilo que realmente se verifica quanto ao modo como operam as nossas convenções, a mesma definição nada altera quanto à exigência de conhecimento mútuo. Assim, seja qual for o número de membros de P que mantém a convenção operacional, as cláusulas que a definem como tal têm de continuar a ser, entre eles, conhecimento comum, mantendo-se assim, para populações suficientemente grandes, a dificuldade identificada por Binmore.

3. Conhecimento comum de convenções alternativas

As dificuldades que a literatura sobre o tema tem vindo a apontar a respeito da possibilidade do conhecimento comum tem igualmente sugerido a alguns autores que o instrumental analítico da teoria dos jogos não é suficiente para captar todas as nuances de um fenómeno social tão complexo como a convenção; ou, pelo menos, não a teoria dos jogos na versão herdada de Von Neumann/Morgenstern, que pode ser designada por racionalista, em que são propostos certos axiomas que governariam o comportamento de agentes idealmente racionais. É por estes motivos que a definição de Lewis é criticada por autores como estando dependente de uma visão demasiado intelectualizada do comportamento dos seres humanos reais, propondo uma abordagem alternativa através da teoria dos jogos evolucionária, a qual não exige dos jogadores que estes sejam sequer criaturas pensantes, permitindo assim à teoria abarcar fenómenos de regularidades comportamentais entre, por exemplo, insectos ou peixes. A ideia central desta abordagem é a de que certos mecanismos não-rationais de coordenação são seleccionados no mundo natural através de um processo semelhante ao da selecção natural das espécies, em que o sucesso de uma regularidade comportamental em

então não pode falhar a formação de crenças de 3ª ordem, e assim sucessivamente enquanto existirem premissas auxiliares.

colonizar uma sociedade depende dos benefícios de sobrevivência que este oferece a uma determinada população ou espécie¹⁰.

Tyler Burge¹¹ é outro autor que critica a definição de Lewis como igualmente demasiado intelectualizada, por comparação com o modo como as convenções operam na realidade. O alvo da sua crítica é mais limitado e, mais uma vez, relacionado com o tema do conhecimento comum e com o modo como a definição de Lewis tenta capturar algumas intuições fundamentais acerca das convenções. Temos vindo a tratar de uma delas: a necessidade dos membros de uma população possuírem crenças verdadeiras acerca uns dos outros de modo a poder-se dizer que o seus estados psicológicos correspondem àqueles que intuitivamente atribuímos aos agentes envolvidos em convenções sociais. A outra intuição fundamental, como vimos, é a da arbitrariedade das convenções, perfeitamente capturada pela definição Lewisiana de convenção como selecção de um dos equilíbrios próprios de um problema de coordenação. A definição inclui, portanto, a cláusula segundo a qual é conhecimento comum em P que existe pelo menos uma alternativa R' a R, tal que a crença de que todos preferem fazer R' em vez de R, e esperam que todos os outros façam R', oferece a cada membro de P uma razão decisiva para fazer igualmente R', e tal que não existe normalmente a possibilidade de fazer simultaneamente R e R'.

Burge afirma que esta condição é demasiado exigente e oferece exemplos de convenções operacionais que o demonstram. É perfeitamente possível que tenham existido ao longo da história comunidades linguísticas de tal modo isoladas que desconhecem por completo o uso de outras linguagens diferentes da sua. Burge considera que uma tal situação configuraria um caso de uma convenção cujas partes desconheciam a existência de convenções alternativas. Mais, essa comunidade poderia entrar em contacto ou ser absorvida por uma outra bastante mais alargada, em que vigorasse uma outra convenção, e recusar-se todavia a seguir o interesse comum na reciprocidade, devido, por exemplo, a um apego religioso à sua linguagem, baseado em crenças mítico-mágicas, segundo as quais as palavras da sua linguagem se encontrariam revestidas de um poder sobrenatural. Também neste caso, segundo Burge, uma das condições da definição não é cumprida, pois não se passaria a verificar uma preferência condicional pela alternativa exibida pela maioria. Ou seja, a preferência geral por uma alternativa não ofereceria aos membros da população primitiva uma razão decisiva – motivacionalmente suficiente, – para modificar a sua conformidade para com uma regularidade comportamental.

Um outro exemplo aponta para a descoberta de geometrias alternativas à de Euclides, supostamente passíveis de mapear o espaço físico de forma igualmente eficaz. Acontece que durante muito tempo se pensou que estas não existissem, o que tornaria o uso da geometria euclidiana supostamente não-convencional. Contudo, este caso não é estruturalmente diferente do da comunidade linguística primitiva, diferindo dele apenas quanto à sofisticação daquilo que seria ou não objecto de conhecimento comum. Face a isto, estaríamos dispostos a considerar que somente depois de se ter feito a descoberta é que o uso de uma geometria euclidiana passou a ser convencional? Burge conclui, assim, que o carácter arbitrário das convenções não se encontra relacionado com quaisquer factos relacionados com os estados mentais ou

¹⁰ K. BINMORE, op. cit., p. 18, faz notar que no tratamento de Schelling a justiça de um certo equilíbrio pode constituir um ponto focal que determina a sua selecção. Neste caso, estaríamos também perante um equilíbrio selecionado por um factor que não se encontra capturado pelo tipo de racionalidade envolvido na tradicional teoria dos jogos.

¹¹ Ver o seu «On Knowledge and Convention», *The Philosophical Review*, 84, 2 (1975) 249-255.

psicológicos das partes, i.e., com aquilo que é conhecimento comum numa população, mas sim com aquilo que se designaria como a «lógica própria da situação», a qual dependeria essencialmente do acidente histórico. A opção por uma regularidade em detrimento de outra, ou até uma certa rigidez conservadora no apego a certas convenções – como, por exemplo, a insistência de um idoso em retirar o chapéu ou fazer vénias às senhoras – dependeria, assim, do facto de essa convenção preencher uma função social mais ou menos importante para uma determinada comunidade ou um sector da população, dependendo esse juízo exclusivamente dessa comunidade e desse sector da população. Burge acredita, portanto, que o conceito de função social é mais adequado para explicar a intuição relacionada com a arbitrariedade das convenções, nomeadamente que essa arbitrariedade reside no facto de as convenções não dependerem de quaisquer leis naturais, psicológicas ou metafísicas, consistindo apenas em criações humanas, resultado de acidentes históricos – o que se adequa, de certo modo, ao conceito de saliência -, preenchendo funções sociais contingentes de população para população. A teoria da propriedade de Hume parece-me ser um exemplo deste tipo de mecanismo, em que as convenções relativas à propriedade, não se podendo considerar verdadeiras ou falsas, na sua qualidade de criações humanas, tornam certos factos verdadeiros ou falsos, tais como, por exemplo, o facto de esta casa ser minha.

4. Avaliação das críticas à definição Lewisiana de convenção

O próprio Lewis veio mais tarde a admitir, provavelmente face ao tratamento sofisticado que a sua formulação original de conhecimento comum veio a receber, que o uso do termo foi talvez «infeliz», dado não existir qualquer garantia de que venha realmente a tornar-se conhecimento ou, sequer, que este venha a ser verdadeiro¹². Esta questão já foi abordada e já vimos que seria mais apropriado falar-se de razão comum para acreditar, ou de potencial crença comum, num certo evento ou proposição.

As críticas de Binmore e Burge são, de certa maneira, previsíveis no sentido em que apontam para o facto de a possibilidade de se alcançar algo como conhecimento ou crença potencial comum depender fortemente do contexto em causa. Como vimos, existem certas situações bem delimitadas, principalmente entre uma população com um número reduzido de membros, em que é plausível considerar-se a possibilidade de se vir a alcançar uma situação de, pelo menos, crença potencial comum¹³. Neste aspecto, as várias contingências contextualmente dependentes que impedem que isso venha a acontecer não parecem diferir das contingências que normalmente interferem com a aquisição do conhecimento ou crença individual, tais como a distração, erro perceptivo, ausência de competências intelectuais adequadas, alucinações, etc. Todavia, mantém-se a questão de saber como pode uma convenção adoptada por milhões de pessoas – como a de conduzir pela direita - depender, segundo a definição, de um truísmo, i.e., de um evento E, tal que E não pode ocorrer sem que (CK)E.

¹² Ver D. LEWIS, «Truth in Fiction», *American Philosophical Quarterly*, 15 (1978) 44, n. 13.

¹³ Suponho que um determinado indivíduo tenha sido o primeiro jogador de futebol a dedicar um golo ao filho recém-nascido executando o gesto de balouçar uma criança nos braços em frente a câmaras de televisão para milhões de espectadores (talvez Bebeto no Mundial de 94 nos Estados Unidos). Este exemplo permite-nos considerar a possibilidade de nos dias de hoje poderem ocorrer eventos tais que não podem ocorrer sem que se tornem potencial crença comum numa população bastante alargada.

É curioso que a história nos ofereça um exemplo de um evento desse tipo quando em 1793, na Dinamarca, foi decretado por lei – por acordo explícito, portanto – que a circulação rodoviária, ao contrário do que até aí acontecia, se passaria doravante a fazer pela direita. É certo que não foi o caso que todos os cidadãos da Dinamarca se tivessem observado uns aos outros a tomarem conhecimento deste decreto. Contudo, é suficiente que cada um deles saiba que tratando-se de uma lei do Estado é natural que, de um modo ou de outro, todos acabem por vir a saber. De início poderão ter-se dado alguns acidentes, devido a uma não-total conformidade para com a nova regularidade, i.e, o grau de convencionalidade, de acordo com a definição de Lewis, poderá ter sido baixo ao início e ter vindo progressivamente a aumentar até se atingir um grau de quase total conformidade. Trata-se, certamente, de um exemplo em que as insignificantes exceções, tanto ao sistema de expectativas mútuas como ao sistema de preferências condicionais que definem a convenção, não passarão de bizarras temporárias – corridas suicidas, enganos relacionados com o consumo de álcool ou idade avançada, etc. Seria implausível acreditar que uma regularidade como esta, em que da conformidade geral para com a mesma dependem diariamente as nossas vidas, não fosse conhecimento comum – na sua forma qualificada - numa determinada população¹⁴. Por outro lado, convenções existentes em populações bastante mais pequenas – como a população académica mundial de lógicos - parecem ser também passíveis de se tornarem conhecimento mútuo com alguma facilidade, dado o quão exigentes são as condições de acesso, para os membros dessa população, quanto à sofisticação dos conhecimentos que é necessário possuir-se – poderá, talvez, ser inconcebível que um membro dessa população não tenha lido um determinado livro ou que não leia com regularidade uma determinada publicação periódica.

O estudo das convenções pode, assim, suscitar duas questões diferentes a seu respeito, uma de natureza estática, se quisermos, e outra de natureza dinâmica. A primeira consiste na tentativa de explicar de que modo, ou através de que mecanismos, têm origem as convenções; a segunda tenta, por seu lado, descobrir os mecanismos através dos quais uma convenção se mantém operacional ao longo do tempo e é transmitida culturalmente de geração para geração. À primeira vista, a análise de Lewis parece encontrar-se mais apta para responder de forma perfeitamente satisfatória à primeira destas questões. É certo que Lewis tem também algo a dizer a respeito da segunda, pois uma resposta imediata a este respeito fará notar que uma convenção se mantém operacional devido à existência numa população de um sistema de expectativas mútuas e de um sistema de preferências condicionais. Contudo, isto não é tudo quanto se pode dizer a respeito da transmissão e durabilidade de uma convenção.

Voltando à primeira questão, parece-me difícil que uma convenção tenha origem sem qualquer tipo de crença ou conhecimento mútuo, pois a própria natureza do mecanismo de selecção – a percepção mútua de algo saliente – não pode deixar de o exigir. É possível que dois remadores iniciantes atinjam um equilíbrio de coordenação de forma instintiva, sem pensarem muito no assunto, cada um deles concentrando-se apenas em remar ao ritmo do

¹⁴ D. MONDERER e D. SAMET, em «Approximating Common Knowledge with common beliefs», *Games and Economic Behaviour*, 1 (1989) 170-190, propõem um enfraquecimento da condição do conhecimento comum, segundo a qual, para que uma convenção se mantenha operacional numa população, é suficiente que o evento público E – o truísmo -, seja já não conhecimento comum, mas apenas acreditado com um certo grau de probabilidade p , ficando por determinar, consoante as circunstâncias de cada situação, acima de que grau a convenção se consegue manter operacional e o grau abaixo do qual se dá o contrário.

seu companheiro, sem que seja necessário possuírem um sistema de expectativas mútuas. Mas isto é o que a própria explicação nos indica: que é possível alcançar-se, por mero acaso, uma situação de equilíbrio, sem que esse equilíbrio tenha sido voluntariamente procurado pelas partes. É possível que os remadores continuem a remar durante léguas acreditando que apenas um deles é responsável pela coordenação. Mas estaríamos dispostos a chamar a isto uma convenção? Esta ausência de crenças verdadeiras acerca das expectativas e crenças dos outros não contraria precisamente aquela que é uma das nossas intuições fundamentais acerca de convenções?

Um grupo de caçadores-recolectores decide ir caçar coelhos após ter montado um acampamento para passar a noite. Antes de se separem, cada um para seu lado, surge do meio da floresta um veado. Espontaneamente, sem qualquer acordo explícito, os caçadores rodeiam-no e abatem-no com os instrumentos que têm à disposição. Nessa noite alimentam-se como até então nunca o tinham feito. A partir desse dia passam a caçar veados em vez de coelhos. Com o decorrer do tempo os pais vão ensinando os filhos a caçar veados. O território que habitam encontra-se densamente povoado por veados e outra caça grossa, e pode mesmo acontecer que as gerações futuras desenvolvam um mito acerca de coelhos, em que estes se tornam criaturas sagradas, e de acordo com o qual foi um coelho ancestral que ensinou os homens a caçar veados. Este exemplo não me parece estruturalmente diferente do de Burge acerca da comunidade linguística isolada; aliás, o grau de sofisticação em causa das duas comunidades parece ser bastante semelhante. Se no caso da linguagem é sempre mais complicado imaginar-se uma situação histórica actual em que certos indivíduos alcançam pela primeira vez um determinado equilíbrio de coordenação de tipo linguístico, numa situação nova para eles, em que ambos sabem – nem que seja potencialmente - que existem equilíbrios linguísticos alternativos, tal possibilidade não é, contudo, de todo implausível. Mais uma vez, o conhecimento comum de convenções alternativas é algo que se encontra dependente do contexto e que pode variar em grau de uma situação para outra. O que não parece possível é que, na sua origem, uma convenção possa ter-se estabelecido sem a formação de um sistema de expectativas mútuas, o qual depende necessariamente da formação de expectativas ou crenças de ordem-n ou, o que é o mesmo, de conhecimento comum ou potencial crença comum - como lhe quisermos chamar.

De que modo podemos, então, encarar os exemplos de Burge que pretendem eliminar da definição de convenção a cláusula do conhecimento mútuo de alternativas? Uma das maneiras é conceder que factores como o hábito, as circunstâncias específicas ou a formação de crenças acerca das convenções, que as cristalizam e as transformam em factos naturais ou metafísicos, acabam por ser decisivas na preservação das convenções ao longo das gerações, tendo um papel decisivo para que estas se mantenham operacionais. Concluindo-se, assim, que o conceito de função social de uma convenção possui uma capacidade explicatória que a noção de problema de coordenação não possui quando se tenta responder à questão dinâmica acerca de como se mantêm operacionais as convenções ao longo do tempo, eliminando-se, assim, essa cláusula da definição geral.

A meu ver, não me parece que venha grande mal ao mundo se nos recusarmos a conceder o estatuto de convenções aos exemplos de Burge, afirmando que fenómenos que começaram como convenções, a certa altura deixaram de o ser – ou vice-versa. Tal parece-me uma questão de terminologia que podemos resolver estabelecendo uma convenção. Penso, aliás, que a análise de Lewis é suficientemente rica para nos oferecer os elementos de análise que nos permitem alcançar uma explicação para os fenómenos revelados nos exemplos de Burge. Consideremos o efeito do mecanismo da promessa na interação social.

Como vimos, um problema de coordenação caracteriza-se por ter mais do que um equilíbrio próprio. Ao estabelecer-se uma convenção um desses equilíbrios é selecionado e passa a verificar-se uma regularidade comportamental baseada nesse equilíbrio. No caso do acordo explícito entre dois amigos que combinam voltar a encontrar-se num determinado local, pode suceder que um dos seus encontros futuros seja de tal modo importante que ambos fazem promessas solenes de não faltarem ao próximo encontro. Estaremos, ainda assim, perante um problema de coordenação? Se considerarmos uma matriz com nove entradas, com três equilíbrios próprios na diagonal, como representado o problema de coordenação original destes dois indivíduos, podemos afirmar que a promessa alterou os «payoffs» da entrada que correspondia ao equilíbrio por eles selecionado, aumentando os valores em relação aos dois outros equilíbrios possíveis. Tendo em conta que passa a existir apenas um equilíbrio próprio na matriz, teremos de concluir que deixou de existir um problema de coordenação. O que sucedeu foi que o desejo de ambos de cumprirem a promessa, fortalecido pela pressão e censura social envolvidas, determina que apenas um dos três equilíbrios iniciais é, de um ponto de vista racional, viável para as partes envolvidas.

Do mesmo modo, outros mecanismos sociais que não as promessas operam de modo a conferir às convenções essa função social que é tão determinante para a sua manutenção, atribuindo-lhes um valor simbólico tal que a hipótese de não-conformação ou a adopção de outras convenções passa a ser excluída, devido ao risco de instabilidade social que isso acarretaria. A análise de Lewis através do instrumentário técnico da teoria dos jogos permite, assim, compreender o apego conservador ou o tradicionalismo a respeito de certas formas de comportamento. Aliás, ele próprio reconhece que todas as nossas convenções se encontram destinadas a transformarem-se em normas, pois o sistema de expectativas mútuas e o sistema de preferências condicionais tende a suscitar reações desfavoráveis para com qualquer membro da população que viole a convenção, contribuindo assim para fortalecer a preferência condicional de cada membro em cumprir a convenção ou a norma.

Em suma, e quanto à questão estática, conclui-se que é necessário existir conhecimento comum – qualificado da maneira já referida – para que uma convenção se estabeleça e, ainda que implicitamente (ou potencialmente, no sentido de que cada um dispõe da evidência para chegar a essa conclusão), as partes que se encontram perto, no tempo e no espaço, da origem da convenção, conhecerão inevitavelmente a existência de convenções alternativas. Quanto à questão dinâmica, conclui-se que provavelmente muitas das convenções se transmitem culturalmente de geração para geração pelo facto de se transformarem em normas e se mantêm operacionais já não porque seja conhecimento comum em P que são soluções para problemas de coordenação, mas sim porque adquirem um valor simbólico indispensável para a preservação da estabilidade social, sem que isso afecte o seu carácter convencional original.

De um ponto de vista moral, talvez constitua um facto feliz que o carácter convencional de uma norma, e a existência de alternativas ao equilíbrio original, seja realmente conhecimento comum numa dada população. Não porque seja motivo de cuidado o senhor idoso pensar que a juventude já não tem maneiras porque não retira o chapéu na presença de senhoras, mas sim porque as preferências legítimas dos membros de uma população se vão modificando, e o apego conservador a certas normas convencionais pode prejudicar os direitos de indivíduos e minorias que têm interesse em seguir convenções alternativas.