

ARTICLE

# Frames, reasoning, and the emergence of conventions

Nicola Campigotto 

Department of Economics, University of Verona, via Cantarane 24, 37129 Verona, Italy  
Email: [nicola.campigotto@univr.it](mailto:nicola.campigotto@univr.it)

(Received 06 September 2018; revised 25 August 2019; accepted 08 September 2019; first published online 18 February 2020)

## Abstract

This paper examines the perceptual and reasoning processes that underpin regularities in behaviour. A distinction is made between situations as they are, or as described by an omniscient external observer, and situations as agents see or frame them. Different frames can stem from differences in culture, experience and personality, as well as from other context-specific factors. Drawing upon David Lewis's *Convention* (1969), I show that consistency between reasoning and experience does not preclude individuals from understanding the same state of affairs differently, and that agents' beliefs about others' beliefs may well be wrong. As a result, cases may occur in which conventions are sustained by false but mutually consistent and self-confirming beliefs.

**Keywords:** common reason to believe; conventions; David Lewis; frames; reasoning

## 1. Introduction

Cognitive processes have long been recognized as an important component of social interactions, and their study has provided valuable insights into the nature and causes of human behaviour. Yet social sciences continue to reflect an ambivalent stance towards perceptual and cognitive assumptions. Although most people agree that individuals perceive and interpret the environment within which they act inaccurately, decision and game-theoretic models frequently posit that agents represent situations to themselves in a systematically correct way. This makes subjective appraisals coincide with reality and uncertainty boil down to a probability distribution over a set of well-defined events.

Early attempts by economists to understand how knowledge is acquired and transmitted trace back to the works of Friedrich Hayek (1937, 1952) and Herbert Simon (1955). For Hayek (1937: 36, 46), the information on which an individual bases his decisions is 'things as they are known to (or believed by) him to exist, and not in any sense objective facts'. Thus, the study of social processes 'must necessarily run in terms of assertions about causal connections, about how

experience creates knowledge'. The differences with more recent epistemic frameworks (see e.g. Brandenburger 2014 and the textbooks by Gintis 2009 and Perea 2012) are not merely a matter of jargon. The latter provide powerful tools for representing beliefs, conditional to the occurrence of certain events, in situations that are unambiguously defined, but do not venture to explain how agents form their mental models (Sugden 2011a).

Possibly in response to this, Lewisian models – often advanced to elucidate the functioning of institutions and social norms – have been receiving increasing attention (Aoki 2011, 2017; Sillari 2013; Hédoin 2014, 2017; Hindriks and Guala 2015). All build on David Lewis's *Convention* (1969), which seeks to capture the role of reasoning in the emergence and maintenance of regularities in behaviour. For Lewis, a pattern in the behaviour of a group of individuals is a convention if and only if it is customary, expected and satisfies a mutual optimality condition. Moreover, these properties must be common knowledge (in a sense that will be clarified below) in the group to which the convention applies.

Notwithstanding its importance, the strand of research initiated by Lewis contains an element of ambiguity as to whether – and if so, to what extent – stable conventions require reasoning to be symmetric across individuals. As elaborated in Section 2 below, some scholars have maintained that for a convention to be in place, agents must have common background information and inductive standards; others have suggested that this might not be the case, without further discussing the point. Intuitively, to say that in a certain situation members of a group share the same inductive standards is to say that: (i) they have a common view or understanding of the relevant features of that situation; (ii) taking this view as a premise, they all reach the same conclusion – which, for instance, may consist of a proposition of the kind 'I ought to do  $x$ '; and (iii) they all have reason to believe that others have made the same inference.

Depending on context, the assumption of shared or symmetric reasoning may or may not be appropriate. For instance, consider the following examples.

**Example 1** (Meeting at the bar). Alice and Bob meet accidentally in a bar. For both of them, making the acquaintance of someone unrelated to work is a breath of fresh air. They enjoy their time together, and, without any prior agreement, they start meeting at the bar every week. However, their motives and understanding of the situation differ considerably. Alice sees Bob as a new friend to share stories with, and nothing in Bob's behaviour gives her reason to believe that he may view their relationship any differently. Bob, however, is immediately infatuated with her. Although aware that Alice just sees him as a friend, he hangs on and continues to meet her, hoping that someday she will reciprocate his feelings. After all, Bob thinks, good things come to those who wait . . .

**Example 2** (Delinquency and violence in juvenile gangs). When privately interviewed, gang members often reveal deep uneasiness about their behaviour (Matza 1964). Yet since they do not express their misgivings publicly, they give the impression of being genuinely committed. Each believes that other members confer higher status on those who engage in violent acts, and some actually regard delinquency and violence as the proper way to express their identity. On

the other hand, no one seems to consider the possibility that many of his peers behave violently not for the sake of it, but because they believe that other members believe that gangs should behave that way, they believe that others believe that they themselves believe that gangs should behave that way, and so on. Conformity signals endorsement and confirms the false belief that the group as a whole supports the gang's conduct (Bicchieri and Fukui 1999).

The examples illustrate two related points. First, some components of an individual's view of a situation may not be shared by others. Bob has romantic feelings for Alice and thinks of their meetings as occasions to win her love, whereas Alice only likes Bob as a friend. Similarly, some gang members take pride in their behaviour, while others conform out of a fear of being rejected or punished by the group. Second, and importantly, agents' beliefs about others' motives and reasoning may be imprecise or false. Alice is wrongly convinced that Bob thinks the same about her as she does him, and each gang member wrongly believes that all other members endorse delinquency. In both examples, false beliefs about other people co-exist with adherence to a regularity in behaviour. Clearly, in each case there are *some* beliefs that are correct and common to all agents (for instance, each gang member correctly believes that each other believes that all other members approve of the group's behaviour). Nevertheless, to limit discussion to correct beliefs only, and to think of the above conventions as being underpinned by symmetric reasoning processes, would give a misleading picture.

In this paper I seek to establish weaker, non-symmetric conditions for the emergence of Lewisian conventions. This is done by making a distinction between the world as it is and the ways individuals perceive and interpret it. I draw from two separate bodies of literature, one related to the concept of self-confirming behaviour (see Section 5 and the references therein) and the other concerned with the framing of situations (Bacharach 1993, 2003, 2006; Turner 2001). 'Frames' are bundles of concepts employed in an individual's subjective representation of objects and events. Reasoning processes operate upon these representations, and can vary considerably depending on which concepts are used to frame a situation. Moreover, similarities in agents' frames can be thought of as a stylized fact about their culture, as human mental processes operate over culturally developed assemblies of experience (Bacharach 1993; Gintis 2016).

The paper aims to contribute to the literature in two ways. First, it stresses that for a convention to emerge and persist over time, agents' reasonings need not be symmetric. More specifically, it shows that the development by individuals of consistent higher-order beliefs does not prevent their views of a situation and their standards of inductive inference from differing in important respects from one another. Second, it allows for cases where conventions are based on false but self-confirming beliefs. This requires that the conventional course of behaviour never prove the agents wrong; in the words of Hayek (1937: 39–40), 'the subjective data, given to the different persons, and the individual plans, which necessarily follow from them, are in agreement', and 'there is consequently a conceivable set of external events which will allow all people to carry out their plans and not cause any disappointments'.

The paper unfolds as follows: Section 2 introduces Lewis's theory and discusses the symmetric reasoning assumption. Section 3 proposes a frame-dependent model for the emergence of Lewisian conventions. Section 4 gives an example drawn from history. Section 5 reviews some related literature. Section 6 concludes and suggests possible directions for future research.

## 2. Conventions and reason to believe

To appreciate the depth of Lewis's thought, two points need to be made. First, his analysis was only originally meant to apply to coordination problems. However, in almost no case does Lewis rely on properties of an action other than its being optimal given an agent's beliefs. The same framework can therefore be applied to a much broader set of interactions.

Second, although Lewis is commonly credited with being the first to introduce the concept of common knowledge,<sup>1</sup> his work actually concerns beliefs rather than knowledge (Cubitt and Sugden 2003; Sillari 2005, 2008). More precisely, Lewis's interest lay in assessing the modes of reasoning that underpin beliefs and behaviour. The framework is that of epistemic rationality: by saying that agent *i* has reason to believe proposition *x*, we mean that *i* is justified in believing that *x* is true by virtue of the reasoning he endorses and the evidence he possesses. Lewis eventually admitted that the term 'common knowledge' used in his definition of convention 'was unfortunate, since there is no assurance that it will be knowledge, or even that it will be true' (Lewis 1978: 44, cit. in Sillari 2008). To avoid confusion, hereafter I use the term 'common reason to believe' (CRTB).

Lewis's (1969) definition of CRTB goes as follows. Members of a population *P* have common reason to believe a proposition *x* if and only if some state of affairs *A* holds such that:

- (i) everyone in *P* has reason to believe that *A* holds;
- (ii) *A* indicates to everyone in *P* that everyone in *P* has reason to believe that *A* holds;
- (iii) *A* indicates to everyone in *P* that *x*.

As suggested by Cubitt and Sugden (2003), states of affairs can be thought of as Savage's states of the world – 'description[s] of the world, leaving no relevant aspect undefined' (Savage 1954: 9). When conditions (i)–(iii) are met, *A* is said to be a *basis for common reason to believe*. This, together with what Lewis refers to as 'suitable ancillary premises' about agents' rationality and their beliefs about each other's reasoning, allows the development of higher-order reason to believe.

The mutual consistency of individual expectations is of crucial importance in sustaining conventions. For expectations to be consistent, each individual must

---

<sup>1</sup>That is, the idea that all agents know that the others know that all other agents know . . . and so on *ad infinitum* (Aumann 1976). As a matter of interest, it is worth mentioning that an even earlier definition of common knowledge was given by American sociologist Morris Friedell (1967, 1969). For a discussion of the roles of common knowledge and common belief in the development of game theory, see Perea (2014).

have reason to believe that everyone else will adhere to a regularity in behaviour. If agents have a conditional preference for conformity, then each one is motivated to behave in ways which in turn confirm other people's expectations. The challenge is therefore to identify the conditions that allow the development of mutually consistent beliefs about future behaviour. An obvious way of producing first- and higher-order reason to believe, Lewis observed, consists in explicitly agreeing to stick to an action or course of action. Another important factor influencing beliefs and behaviour is salience, namely the perception that some elements of one's view of a situation stand out from others. Finally, in the case of repeated interactions, precedence may emerge as a special kind of salience, resulting in a tendency to take those actions which have proven successful in the past.<sup>2</sup>

### 2.1. On symmetric reasoning

What individuals have reason to believe depends on their logic of reasoning and the information on which they assess situations. In particular, Lewis's common reason to believe can only obtain when all members of a group have reason to believe that they all share the same inductive standards and background information. It therefore cannot be understood from a purely individualist perspective but requires people to endorse lines of reasoning of the kind 'Since I have reason to believe  $x$ , so may others'.<sup>3</sup>

Lewisian models, however, often go beyond this and assume that agents *actually* share inductive standards. For example, Hédoin (2014) explicitly postulates that individuals commonly know that they share the same modes of reasoning with respect to a given set of situations. Lewis's original framework, too, is generally interpreted as requiring common inductive standards. For Sillari (2005: 391), 'in some cases inductive standards are shared by groups of agents, as, for example, in those cases in which common reason to believe comes about', and Bicchieri (2006: 37) observes that 'Lewis's argument is crucially dependent on assuming shared inductive standards'. Similarly, Vanderschraaf (1998: 362) claims that 'a crucial assumption in Lewis's analysis of common knowledge is that agents know they share the same rationality, inductive standards, and background information with respect to a state of affairs'.

Lewis's definition of CRTB does not necessarily require reasoning to be identical across all agents. For instance, it allows individuals to infer the same conclusion from the same premise through somewhat different paths. Also, it does not mean to suggest that *all* inductive inferences need to be shared; the condition only applies to a particular set of propositions which are relevant to the situation being considered. However, Lewis's definition does require individuals to reason symmetrically: if a proposition ('A holds') indicates another proposition ( $x$ ) to a member of  $P$ , then it does the same to all other members of the population

<sup>2</sup>Sugden (2011b) uses the concept of salience to explain how and why a behavioural regularity may come about. The role of perceptions of similarity with past instances of a decision problem has been discussed, among others, by Gilboa and Schmeidler (1995) and Alberti *et al.* (2012).

<sup>3</sup>Credit for this point goes to Herbert Gintis.

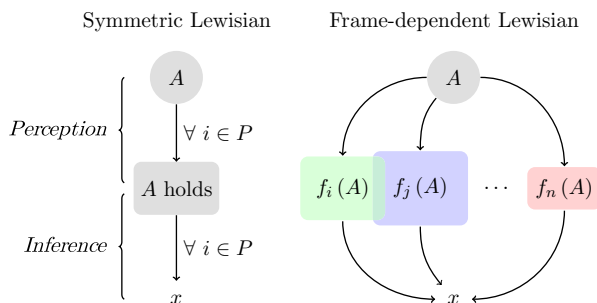


Figure 1. Perception-inference processes.

(Hédoin 2014, 2017; Vanderschraaf and Sillari 2014). This requirement is controversial; Sillari (2008: 31) acknowledges it as being ‘far from innocuous’, and for Aoki (2011: 27) ‘it is admittedly strong and may often fail to apply’.

This symmetry in agents’ reasoning arises in Lewis’s framework as a consequence of the fact that no attempt is made to distinguish between the world *as it is* and the world *as individuals see it*. Lewis’s and Lewisian models implicitly assume that states of affairs are public and perfectly self-revealing, implying that all agents hold the same, correct view of  $A$  (Cubitt and Sugden 2003). If  $A$  obtains, then each agent has reason to believe that  $A$  holds; this, in turn, leads everyone to infer  $x$ . A graphical representation of the perception-inference process underlying Lewis’s CRTB is shown in the left-hand panel of Figure 1 (ignore the right-hand panel for now; we will return to it later).

When is an event public and self-revealing? As an example, consider the coordination problem discussed by Lewis (1969: 52) to introduce conventions by agreement:

[s]uppose the following state of affairs – call it  $A$  – holds: you and I have met, we have been talking together, you must leave before our business is done; so you say you will return to the same place tomorrow. Imagine the case. Clearly, I will expect you to return. You will expect me to expect you to return. I will expect you to expect me to expect you to return.

This state of affairs can reasonably be seen as capable of generating higher-order expectations through roughly common lines of thought: (i) my view of  $A$  is the same as yours (you had to leave and you said you will return to the same place tomorrow), and both you and I have reason to believe that this is indeed the case; (ii)  $A$  indicates to both that both have reason to believe that  $A$  holds; and (iii)  $A$  indicates to both that you will return. Therefore,  $A$  is a basis for CRTB in Lewis’s sense.

In other cases, and especially when individuals are driven by different motives or interpret the same situation differently, the symmetry-in-reasoning condition may not hold. Yet agents may still form consistent higher-order beliefs. Cubitt and Sugden (2003) offer useful insights in this regard by introducing the concept of ‘distributed reason to believe’. In discussing why American drivers drive on the

right side of the road, they observe that no driver can have direct knowledge of other people’s experiences and driving habits, and that each agent’s decisions are made on the basis of experience of a particular sample of American drivers. This makes it natural to think of drivers’ expectations as being licensed by mental state inferences that differ slightly from one another. In a related vein, Cubitt and Sugden (2014) present a model where each agent endorses a private ‘reasoning scheme’. Reasoning schemes are assumed to share a core of ‘common reason’, which consists of those modes of reasoning that are shared across individuals and that each attributes to everyone in the population. However, apart from that, agents’ reasonings may differ to a large extent. The model presented here builds on these ideas by introducing elements of framing theory into Lewis’s framework.

### 3. The model

#### 3.1. Some formal apparatus

##### 3.1.1. Reason to believe and indication

The notation follows that of Cubitt and Sugden (2003) and adapts it to a possible worlds setting. Let  $P = \{1, \dots, n\}$  be a finite population of agents and let  $\mathcal{A}$  be a non-empty set of possible worlds or states of affairs. A proposition is a subset  $x \subseteq \mathcal{A}$ , and the set of all possible propositions is given by the power set  $\mathbb{P}(\mathcal{A})$ . If  $A \in \mathcal{A}$  belongs to  $x$ , then the proposition  $x$  is true in world  $A$ . For each agent  $i \in P$ , let  $\mathcal{R}_i$  be a modal operator from  $\mathbb{P}(\mathcal{A})$  to itself.  $\mathcal{R}_i(x)$  contains those worlds in which  $i$  has reason to believe  $x$ . Reason to believe can be nested to arbitrary depths; I write  $\mathcal{R}_i[\mathcal{R}_j(x)]$  to denote the proposition that  $i$  has reason to believe that  $j$  has reason to believe  $x$ , and I write  $\mathcal{R}^P(x)$  to mean that  $\mathcal{R}_i[\mathcal{R}_j[\mathcal{R}_k[\dots[\mathcal{R}_n(x)]\dots]]]$  is true for all finite sequences  $i, j, k, \dots, n \in P$ .

Lewis’s indication relation is captured by a three-place modal operator, *ind*, where  $x \text{ ind}_i y$  is the proposition that if  $i$  has reason to believe  $x$ ,  $i$  thereby has reason to believe  $y$ . To write  $A \in (x \text{ ind}_i y)$  is to say that in the context of state of affairs  $A$ , having reason to believe  $x$  provides  $i$  with reason to believe  $y$ . Put another way, in world  $A$  – and possibly in many or even all other worlds – the logic of reasoning that  $i$  endorses justifies an inference from  $x$  to  $y$ . However, there may exist  $j \neq i \in P$  and  $A' \neq A \in \mathcal{A}$  for which this does not hold, meaning that indication has no universal validity. The indication relation is taken to be reflexive: every proposition  $x$  always indicates to  $i$  that  $x$ , that is  $x \text{ ind}_i x$  holds tautologically.

The following axioms are assumed. Let  $x, y$  and  $z$  be propositions; for any distinct  $i$  and  $j$  in  $P$ :

$$[\mathcal{R}_i(x) \cap (x \text{ ind}_i y)] \subseteq \mathcal{R}_i(y), \tag{1}$$

$$[(x \text{ ind}_i \mathcal{R}_j(y)) \cap \mathcal{R}_i(y \text{ ind}_j z)] \subseteq [x \text{ ind}_i \mathcal{R}_j(z)]. \tag{2}$$

Rule (1) states that if  $x$  indicates  $y$  to agent  $i$ , then reason to believe  $x$  implies reason to believe  $y$ . Rule (2) says that if  $x$  indicates to  $i$  that  $j$  has reason to believe  $y$ , and if  $i$  has reason to believe that  $y$  indicates to  $j$  that  $z$ , then  $x$



indicates to  $i$  that  $j$  has reason to believe  $z$ . For example, let  $x$ ,  $y$  and  $z$  be the propositions ‘the thunder is getting louder’, ‘it may rain soon’, and ‘it is a good idea to take an umbrella when leaving the house’, respectively. According to (2), if the thunder indicates to  $i$  that  $j$  has reason to believe that it may rain soon, and if  $i$  has reason to believe that the possibility of showers indicates to  $j$  that it is a good idea to take an umbrella, then having reason to believe that the thunder is getting louder indicates to  $i$  that  $j$  has reason to believe that they should leave the house with an umbrella.

### 3.1.2. Framing structures

A frame is composed of bundles of concepts that are used in an individual’s contextual representation of a situation. For Turner (2001: 13, 101), ‘basic mental operations operate over ... frames’, and ‘it is natural to assume that decision making in any specific situation will depend on what frames are used by the decision maker as conceptual inputs’. Moreover, interacting agents typically base their decisions on their perceptions and assessments of other people’s views. Based on these observations, I define a framing structure as a tuple:

$$\mathcal{F} := \langle P, \mathcal{A}, \{f_i, f_{ij}\}_{j \neq i \in P} \rangle,$$

where for every  $i \in P$ , the function  $f_i: \mathcal{A} \rightarrow \mathbb{P}(\mathcal{A})$  maps world  $A \in \mathcal{A}$  to a proposition  $f_i(A)$  that represents  $i$ ’s view of  $A$ . This introduces a distinction between the world as seen from the external viewpoint of the modeller,  $A$ , and the world as each agent sees or perceives it,  $f_i(A)$ . Different frames may reflect differences in culture, cognitive ability or experience. In a related fashion, for every pair of individuals  $i$  and  $j$ , the function  $f_{ij}: \mathbb{P}(\mathcal{A}) \rightarrow \mathbb{P}(\mathcal{A})$  maps  $f_i(A)$  to another proposition,  $f_{ij}(f_i(A))$ . The latter represents what  $i$  conjectures  $j$ ’s frame to be, given that  $i$ ’s own frame is  $f_i(A)$ .<sup>4</sup>

Several points are worth making here. First, if  $f_i(A) = f_i(A')$  holds for any distinct  $A, A' \in \mathcal{A}$ , then the two worlds are frame-equivalent, meaning that  $i$  cannot discriminate  $A$  from  $A'$ . Second, for all members of  $P$  to hold the same view of  $A$ , it must be that  $f_i(A) = f_j(A)$  for every  $i$  and  $j$  in  $P$ . Third, in the particular case where  $f_i(A) = f_{ij}(f_i(A)) = A$  for all  $i$  and  $j$ , the frame-dependent model reduces to the symmetric Lewisian model. As we have seen, this may be a strong condition to impose; in general,  $i$ ’s and  $j$ ’s frames of  $A$  will be similar in some (perhaps many) respects and different in others. Likewise,  $i$ ’s frame and  $i$ ’s

<sup>4</sup>The idea that individuals rationalize behaviours by attributing thoughts and beliefs to other people has good support in the literature. In his *Theory of Moral Sentiments*, Adam Smith argues that:

[a]s we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. ... it is by the imagination only that we can form any conception of what are his sensations. (Smith 1759: 34)

Similarly, Lewis (1969: 27) notes that ‘we may acquire ... expectations, or correct or corroborate whatever expectations we already have, by putting ourselves in the other fellow’s shoes, to the best of our ability’. More recent contributions on folk psychology generally refer to this as Simulation Theory (Gordon 1986).



conjecture of  $j$ 's frame will typically be correlated with one another, but may also differ in some important respects. For instance, in our earlier example, some gang members view their group's behaviour as morally wrong but conjecture that other members take pride in violence. Fourth, 'framing is logically prior to believing' (Bacharach 2003: 66): if  $f_i(A)$  and  $f_{ij}(f_i(A))$  do not contain proposition  $y$ , then agent  $i$  cannot have beliefs about  $y$  when representing  $A$  to himself. Moreover,  $i$  remains unaware of this because  $y$  has not occurred to him at all.

Finally, the components of an individual's frame may be seen as being characterized by different degrees of salience and choice relevance. The salience of a proposition is related to the frequency with which that proposition is employed in agents' representations of a situation (Bacharach 2003, 2006), whereas choice relevance refers to the ability to influence decisions (Gold 2012). Both salience and choice relevance are typically time- and context-dependent.

### 3.2. Breaking the symmetry

A more general definition of basis for CRTB can now be given.

**Definition 1** (Basis for common reason to believe). Let  $x$  be a proposition. A state of affairs  $A$  is a basis for common reason to believe in  $P$  that  $x$  if and only if:

$$\forall i \in P : A \in \mathcal{R}_i[f_i(A)], \tag{3}$$

$$\forall i \in P : A \in [f_i(A) \text{ ind}_i x], \tag{4}$$

$$\forall i, j \in P : A \in [f_i(A) \text{ ind}_i \mathcal{R}_j[f_{ij}(f_i(A))]]. \tag{5}$$

Condition (3) states that each  $i \in P$  frames  $A$  as  $f_i(A)$  and has reason to believe that  $f_i(A)$  holds. Thus,  $A$  is neither perfectly nor identically self-revealing. Note that it may be that  $A \notin f_i(A)$  (so that agent  $i$  completely misperceives  $A$ ) and yet  $A \in \mathcal{R}_i[f_i(A)]$  (meaning that  $i$  has reason to believe in his view of  $A$ ). This is because, for any proposition  $y$ , the reason to believe operator need not satisfy the truth axiom  $\mathcal{R}_i(y) \Rightarrow y$ . Condition (4) says that  $f_i(A)$  indicates to each  $i$  that  $x$ , i.e. that agents' modes of reasoning license an inference from their view of  $A$  to  $x$ . The perception-inference process corresponding to conditions (3) and (4) is represented in the right-hand panel of Figure 1. If condition (5) holds for some  $A$ , then for each pair of individuals  $i, j \in P$ , having reason to believe  $f_i(A)$  indicates to  $i$  that  $j$  has reason to believe  $f_{ij}(f_i(A))$  – the latter being  $i$ 's conjecture of  $j$ 's view of  $A$ , which may differ significantly from  $j$ 's actual view.

Sometimes (as in the example of Section 4) agents might have reason to believe that other members of  $P$  have the same view of  $A$  as themselves, or at least that they do so in all relevant respects. If this is the case, then  $f_i(A) = f_{ij}(f_i(A))$  for all  $i$  and  $j$ , and (5) reduces to:

$$\forall i, j \in P : A \in [f_i(A) \text{ ind}_i \mathcal{R}_j[f_i(A)]], \tag{5'}$$

which reflects a mutual ascribing of common views. The condition in (5') states that  $f_i(A)$  indicates to each  $i$  that others have reason to believe  $f_i(A)$  as well. Again, this by no means implies an actual commonality of views.

Our ancillary condition for the development of common reason to believe is given in the definition below.

**Definition 2** (Ascription of mutually consistent inductive standards). Let  $y$  be a proposition. In state of affairs  $A$ , members of  $P$  ascribe mutually consistent inductive standards to each other if:

$$\forall i, j \in P, \forall y: A \in [f_i(A) \text{ ind}_i y] \Rightarrow A \in \mathcal{R}_i[f_{ij}(f_i(A)) \text{ ind}_j y]. \quad (6)$$

The intuition behind (6) is that although recognizing that other agents' views differ from his own, an individual may have reason to believe that these views all justify an inference to proposition  $y$ . For instance, in Example 1 earlier: (i) Bob's view indicates to him that meetings with Alice are not going to stop any time soon; (ii) Bob is aware that Alice sees him as a friend, i.e. that Alice's view differs from his own; and yet (iii) Bob has reason to believe that Alice's view (as he believes it to be) indicates to her that the meetings are going to continue. More precisely, to say that (6) holds for some  $A$  is to say that, for all  $i$  and  $j$  in  $P$ , if  $f_i(A)$  indicates any proposition  $y$  to  $i$ , then  $i$  has reason to believe that  $j$ 's view of  $A$  (as  $i$  conjectures it to be!) indicates to  $j$  that  $y$ . As in Cubitt and Sugden (2003: 189, 192), this condition is stronger than is strictly necessary but makes it possible to keep technicalities from dominating the exposition. In the following, I do not require the material implication to hold for all  $y$ s but only for a limited set of propositions required for the development of higher-order reason to believe.

Definition 2 encompasses cases in which agents have reason to believe that others share the same inductive standards, that is, in which:

$$\forall i, j \in P, \forall y: A \in [f_i(A) \text{ ind}_i y] \Rightarrow A \in \mathcal{R}_i[f_i(A) \text{ ind}_j y]. \quad (6')$$

The condition in (6') states that if  $f_i(A)$  indicates any proposition  $y$  to  $i$ , then  $i$  has reason to believe the same to hold true for every other person in  $P$ . This corresponds to Lewis's 'suitable premise' about agents' inductive standards, and dovetails with a condition given by Perea (2007) to discuss the epistemic foundations of Nash equilibrium. As Perea shows, in games with at least three players, Nash equilibrium requires each player to have 'projective beliefs'. Simply put, if  $i$  has projective beliefs, then his belief about  $j$ 's behaviour is the same as his belief about  $k$ 's belief about  $j$ 's behaviour.

The conditions in the above definitions, together with axioms (1) and (2), yield the following proposition, which is a frame-dependent version of Lewis's main result. If a state of affairs  $A$  is a basis for common reason to believe in  $P$  that  $x$ , and if members of  $P$  ascribe mutually consistent inductive standards to each other, then each member of the population can form higher-order beliefs about  $x$ .

**Proposition.** *If the conditions in Definitions 1 and 2 hold for some  $A$ , then  $A \in \mathcal{R}^P(x)$ .*

*Proof.* The proposition follows immediately from applying the definitions to the chain of logical implications used by Cubitt and Sugden (2003: app. 1) to formalize Lewis's CRTB.

$$\forall i \in P: A \in \mathcal{R}_i[f_i(A)] \quad \text{from (3)} \quad (\text{P1.1})$$

$$\forall i, j \in P: A \in [f_i(A) \text{ ind}_i \mathcal{R}_j[f_{ij}(f_i(A))]] \quad \text{from (5)} \quad (\text{P1.2})$$

$$\forall i \in P: A \in [f_i(A) \text{ ind}_i x] \quad \text{from (4)} \quad (\text{P1.3})$$

$$\forall i \in P: A \in \mathcal{R}_i(x) \quad \text{from (P1.1), (P1.3) and (1)} \quad (\text{P1.4})$$

$$\forall i, j \in P: A \in \mathcal{R}_i[f_{ij}(f_i(A)) \text{ ind}_j x] \quad \text{from (P1.3) and (6)} \quad (\text{P1.5})$$

$$\forall i, j \in P: A \in [f_i(A) \text{ ind}_i \mathcal{R}_j(x)] \quad \text{from (P1.2), (P1.5) and (2)} \quad (\text{P1.6})$$

$$\forall i, j \in P: A \in \mathcal{R}_i[\mathcal{R}_j(x)] \quad \text{from (P1.1), (P1.6) and (1)} \quad (\text{P1.7})$$

$$\forall i, j, k \in P: A \in \mathcal{R}_i[f_{ij}(f_i(A)) \text{ ind}_j \mathcal{R}_k(x)] \quad \text{from (P1.6) and (6)} \quad (\text{P1.8})$$

$$\forall i, j, k \in P: A \in [f_i(A) \text{ ind}_i \mathcal{R}_j[\mathcal{R}_k(x)]] \quad \text{from (P1.2), (P1.8) and (2)} \quad (\text{P1.9})$$

$$\forall i, j, k \in P: A \in \mathcal{R}_i[\mathcal{R}_j[\mathcal{R}_k(x)]] \quad \text{from (P1.1), (P1.9) and (1)} \quad (\text{P1.10})$$

... and so on, up to  $A \in \mathcal{R}^P(x)$ .

Line (P1.8) follows from the fact that (P1.6) holds for any pair of individuals in  $P$ . Thus, having reason to believe  $f_i(A)$  indicates to  $i$  that  $k$  has reason to believe  $x$ . Lines (P1.4), (P1.7) and (P1.10), together with the subsequent steps, establish the result.  $\square$

The proposition relaxes Lewis's theorem by providing weaker sufficient conditions for the development of higher-order reason to believe. The distinction between a state of affairs, an individual's private view, and an individual's beliefs about other agents' views allows us to distinguish between two kinds of heterogeneity in reasoning. First, different agents may frame  $A$  in ways that differ considerably from one another, and yet all infer  $x$ . Second, individuals may have reason to believe that others' frames and inferences differ from their own. Furthermore, the beliefs that agents have about others' beliefs may be at odds with those agents' actual beliefs. When this is so, then as long as the course of action is consistent with their reasoning, agents do not receive any evidence that their views are inaccurate and their beliefs false. The following example illustrates this point.

#### 4. An extended example

The understanding of the American Revolution put forward by Jack Rakove, Barry Weingast and co-authors (Rakove 1996; Rakove *et al.* 2005; de Figueiredo *et al.* 2006) has some interesting similarities with the framework developed here. As historians have long recognized, the revolutionary crisis was 'the product of ... dimly perceived and rapidly changing thoughts and situations', and its escalation 'becomes comprehensible only when the mental framework ... into which the Americans fitted the events of the 1760s and 1770s is known' (Wood 1966: 14, 23). Moreover, the century of Anglo-American cooperation that preceded the 1775–83 war offers an example of regularity in behaviour

that is conventional (i.e. customary, expected by each party, and characterized by a conditional preference for conformity) but cannot be adequately explained by symmetrical Lewisian models.

Underlying cooperation, there were profound disagreements about the nature of the rule of law and the division of authority between metropole and colonies, of which neither Americans nor British were fully aware (Reid 1976, 1995; Greene 1986; Zuckert 2005). Long-established practices had been based on contrasting interpretations of the American position within the British imperial order, which did not come to light until Britain's parliamentary intervention in colonial affairs became a matter of paramount importance.

The colonists understood the constitution of the empire as a system of common law rules built on precedent. Between the late 17th and the mid-18th centuries, American control over domestic affairs had continued to grow towards what Greene (1991) has defined a 'negotiated authority': the motherland had retained power over matters related to security and international trade, while colonies had come to administer internal affairs that ranged from religion to taxation and the enforcement of contracts. Moreover, in the eyes of the colonists, well-established customs and self-government practices had to be given a pre-eminent legal status. The Americans 'were espousing the ... theory of a constitution of customary restraints on arbitrary power, a theory that seemed to them to be more compatible with – and more explanatory of – their own constitutional experience both within the colonies and with regard to metropolitan-colonial relationships' (Greene 1986: 71). The fact that the British had long ceased to directly intervene in colonial domestic matters was seen as the foundation for a constitutional claim; sovereignty within the empire, colonists believed, was divided, and the metropole lacked the authority to regulate colonial domestic affairs.

However, on the other side of the Atlantic, a different view of common law constitutionalism had been maturing. By the middle of the 18th century, legislative bodies had considerably extended their control over the executive authority of the Crown. The King-in-Parliament principle had been firmly established, and parliamentary sovereignty had emerged as the core of the British constitution. The Houses had been vested with the power to create, amend or abolish any law. Furthermore, the British saw the division of authority between motherland and colonies as motivated by mere expediency. For them,

the structure of the empire was a matter of policy choice having no constitutional status. Allowing the colonists control over their domestic affairs was a privilege, not a right. ... Moreover, the British had no reason to believe that the colonists saw these issues any differently. (de Figueiredo *et al.* 2006: 389)

The British view had therefore little to do with the American ideas of shared sovereignty and self-government practices as acquired rights which restricted parliamentary authority. British policymakers saw sovereignty as indivisible, and the Houses as having supreme legal power in both motherland and colonial territories.

For decades, Anglo-American relations remained characterized by a mutual ignorance of these different, and difficult to reconcile, views. This caused both parties to rely on incorrect but self-confirming beliefs. For the British, what underpinned cooperation was the threat posed by France, both in Europe and

		Colonies	
		<i>c</i>	<i>d</i>
Metropole	<i>C</i>	<i>B</i> ; <i>b</i>	<i>W</i> ; <i>s</i>
	<i>D</i>	<i>S</i> ; <i>w</i>	<i>T</i> ; <i>t</i>

Figure 2. The Anglo-American cooperation game.

overseas. The metropole's accommodating stance towards the American push for autonomy was largely motivated by the need to maintain a solid front against the French. The colonists feared the French and their native allies, too, but an equally compelling reason to cooperate was provided by their being fine with the political independence they had gained. As North (2005: 1006) put it, 'to the extent that Americans thought about the metropole, it was that Britain was a benign, if remote, presence'. No evidence emerged to reveal that the two parties' constitutional ideas were contradictory and their beliefs inaccurate.

Figure 2 represents Anglo-American cooperation as the equilibrium of a coordination game. Upper and lower case letters represent the choices of, and the payoffs to, the British and the Americans, respectively. *B*, *S*, *T* and *W* stand for best, second best, third best and worst; *C* and *D* denote cooperation and defection. For the British, to cooperate means not intervening in American domestic affairs. For the Americans, it means not revolting against the British. The two parties' preference orderings over the possible outcomes are, respectively,  $(C, c) >_{UK}(D, c) >_{UK}(D, d) >_{UK}(C, d)$  and  $(C, c) >_{US}(C, d) >_{US}(D, d) >_{US}(D, c)$ . The game has two Nash equilibria in pure strategies,  $(C, c)$  and  $(D, d)$ .

This representation reveals very little about players' views and (mis)conjectures about the opponent's view. To apply our framework, let *A* denote Anglo-American relations (as described above and as seen by an omniscient external observer) at some point in time between the 1690s and the 1760s, and let  $f_{US}(A)$  and  $f_{UK}(A)$  denote the contextual features of *A* that the Americans and the British considered relevant. For ease of exposition, and coherently with Rakove *et al.*'s narrative, suppose that conditions (5') and (6') hold, reflecting that neither party had reason to expect the other to view *A* in a significantly different manner from himself.<sup>5</sup> The American and British frames of *A* can be thought of as:

$$f_{US}(A) = \{ \text{common law constitutionalism built on precedent;} \\ \text{colonial self-governance as an acquired right;} \\ \text{negotiated authority as a satisfying status quo} \},$$

$$f_{UK}(A) = \{ \text{parliamentary supremacy;} \\ \text{colonial self-governance as a matter of policy choice;} \\ \text{negotiated authority necessary under the French threat} \}.$$

<sup>5</sup>Recall that (5') and (6') are special cases of (5) and (6), respectively, and are not necessary for the proposition in Section 3 to hold.

**Table 1.** The reasoning schemes underpinning Anglo-American relations

	British point of view	American point of view
(3)	$A \in \mathcal{R}_{UK} [f_{UK}(A)]$	$A \in \mathcal{R}_{US} [f_{US}(A)]$
(4)	$A \in [f_{UK}(A) \text{ ind}_{UK} x]$	$A \in [f_{US}(A) \text{ ind}_{US} x]$
(5')	$A \in [f_{UK}(A) \text{ ind}_{UK} \mathcal{R}_{US} [f_{UK}(A)]]$	$A \in [f_{US}(A) \text{ ind}_{US} \mathcal{R}_{UK} [f_{US}(A)]]$
(6')	$A \in \mathcal{R}_{UK} [f_{UK}(A) \text{ ind}_{US} x]$	$A \in \mathcal{R}_{US} [f_{US}(A) \text{ ind}_{UK} x]$

Finally, let  $x$  be the proposition ‘the existing status quo should be maintained’. Table 1 summarizes the modes of reasoning that underlay the two parties’ behaviour. The first three lines of the table show that despite the significant differences in the British and American views of  $A$ , the latter represented a basis for common reason to believe  $x$ . Line four follows from taking (6’) as the ancillary condition allowing higher-order beliefs to come about.

France’s defeat in the Seven Years’ War was a major environmental shock that produced a change in the two parties’ willingness to cooperate and eventually led to open conflict. For the first time in years, Britain sought to directly intervene in colonial domestic affairs in order to cover the costs of war. Most laws, such as the 1764 Sugar Act and the 1765 Stamp Act, introduced relatively small taxes. Yet they astounded the colonists because they were inconsistent with their long-held beliefs. The American view of common law constitutionalism made the colonists interpret the British interventions as unlawful. The British, in turn, interpreted the American discontent as fading loyalty towards the empire, and reacted by suspending the New York Colonial Assembly. As events unfolded and the revolutionary crisis escalated, the cooperative equilibrium rapidly fell apart.

## 5. Related literature

The case of pre-1760s Anglo-American relations can be recognized as an example of ‘spurious unanimity’, namely a situation characterized by unanimity of preferences without unanimous reasons (Mongin 2016). It is also a case of self-confirming behaviour, a concept introduced into the economics literature by the aforementioned work of Hayek (1937: 51): ‘we may ... very well have a position of equilibrium only because some people have no chance of learning about facts which, if they knew them, would induce them to alter their plans’. In this same line of thought we have Hahn (1977), who describes an economy in which agents formulate conjectures about market conditions from the private signals they receive. Their hypotheses may be imprecise and yet lead to a ‘conjectural equilibrium’, i.e. a set of mutually consistent signals and individual actions that confirm and induce each other.

In game theory, the concept of self-confirming equilibrium was introduced by Fudenberg and Levine (1993) as a coarsening of Nash equilibrium such that no agent ever observes an action that contradicts his beliefs – which need not be correct on off-equilibrium paths. Battigalli *et al.* (2015) study self-confirming equilibria in situations where agents do not know the probability model underlying the variables affecting their choices. If the game is played repeatedly in a stationary environment,

then agents can learn the distribution of payoffs associated with observed strategy profiles. However, ambiguity about unchosen strategies persists and makes them less appealing. Schipper (2018) introduces a notion of self-confirming equilibrium that applies to games with unawareness, that is games where individuals are not aware of some of the actions that can be taken by others. He shows that rational play can enrich agents' information sets, and characterizes the game as the endogenous result of strategic interaction and learning. An epistemic game-theoretic approach to reasoning in games with unawareness has been proposed by Perea (2017), who studies the concept of common belief in rationality in that setting.

The role of agents' views is also central to the analysis of mutually acceptable behaviours by Greenberg *et al.* (2009). The authors show that, even if agents view the same interaction problem as different extensive form games, they may follow a common course of action. Kaneko and Atsui (1999) and Kaneko and Kline (2008) propose an inductive framework to explain how individuals develop and modify their views of the strategic environment. In their models, agents do not have *a priori* knowledge of the structure of the game they play, and gather information by means of occasional random moves.

Another means of studying conventions is through evolutionary game-theoretic models, which rely on random perturbations to introduce variations in the frequencies with which strategies are played by individuals in a population (e.g. Young 1998; Skyrms 2014). Information processing is assumed to be costly, and individuals myopically adapt to the behaviour of other agents. Models of this kind allow one to investigate the roles of mutation and adaptation in determining long-run behavioural patterns, but also result in a downplaying of reasoning. This makes the evolutionary approach a valuable complement to Lewisian models of behaviour, but not a substitute for them.

## 6. Concluding remarks

Ideas from psychology and cognitive sciences have recently started to permeate theories of decision making, resulting in a variety of models that emphasize the role of mental states as drivers of behaviour. In a related field of inquiry, Lewisian models have been investigating how experience and reasoning can allow behavioural regularities to emerge.

Building on an ambiguity in previous research about whether or not conventions require symmetric reasoning, this paper sought to expand Lewis's framework by incorporating the concept of 'frames' in it. I stressed that different agents can understand the same situation differently, and that individuals may believe others' beliefs to differ from their own. This diversity pervades our lives, as shown by examples ranging from day-to-day interactions to deviant behaviours and international relations. The main contribution of the paper is to show that individuals can develop consistent higher-order expectations even when their modes of reasoning differ substantially from one another. It also shows that beliefs about other people's views and inferences may not be true, and that self-confirming patterns of behaviour may emerge that are sustained by false beliefs.

Needless to say, many of these points would benefit from further study. For example, the topological properties of framing structures have largely been left



unspecified, and a formal definition of when agents perceive a particular feature of a situation as salient or choice-relevant is still missing. Finally, embedding the model in a dynamic framework would help explain how the revision of frames and beliefs works. Research along these lines could provide important new insights into the reasoning processes that motivate behaviour.

**Acknowledgements.** I thank Paul Anand, Ennio Bilancini, Samuel Bowles, Franz Dietrich, Nicola Dimitri, János Flesch, Francesco Guala, Andrés Perea, Robert Sugden and two anonymous referees for helpful comments and suggestions, many of which were incorporated in the paper.

## References

- Alberti F., R. Sugden and K. Tsutsui** 2012. Salience as an emergent property. *Journal of Economic Behavior and Organization* **82**, 379–394.
- Aoki M.** 2011. Institutions as cognitive media between strategic interactions and individual beliefs. *Journal of Economic Behavior and Organization* **79**, 20–34.
- Aoki M.** 2017. Strategies and public propositions in games of institutional change: comparative historical cases. *Journal of Comparative Economics* **45**, 171–187.
- Aumann R.J.** 1976. Agreeing to disagree. *Annals of Statistics* **4**, 1236–1239.
- Bacharach M.** 1993. Variable universe games. In *Frontiers of Game Theory*, ed. K. Binmore, A. Kirman and P. Tani, 255–275. Cambridge, MA: MIT Press.
- Bacharach M.** 2003. Framing and cognition in economics: the bad news and the good. In *Cognitive Processes and Economic Behaviour*, ed. N. Dimitri, M. Basili and I. Gilboa, 63–74. London: Routledge.
- Bacharach M.** 2006. *Beyond Individual Choice, Teams and Frames in Game Theory*, ed. N. Gold and R. Sugden. Princeton, NJ: Princeton University Press.
- Battigalli P., S. Cerreia-Vioglio, F. Maccheroni and M. Marinacci** 2015. Self-confirming equilibrium and model uncertainty. *American Economic Review* **105**, 646–677.
- Bicchieri C.** 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
- Bicchieri C. and Y. Fukui** 1999. The great illusion: ignorance, informational cascades, and the persistence of unpopular norms. *Business Ethics Quarterly* **9**, 127–155.
- Brandenburger A.** 2014. *The Language of Game Theory: Putting Epistemics into the Mathematics of Games*. Hackensack, NJ: World Scientific.
- Cubitt R.P. and R. Sugden** 2003. Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. *Economics and Philosophy* **19**, 139–146.
- Cubitt R.P. and R. Sugden** 2014. Common reasoning in games: a Lewisian analysis of common knowledge of rationality. *Economics and Philosophy* **30**, 285–329.
- de Figueiredo R.J.P., J.N. Rakove and B.R. Weingast** 2006. Rationality, inaccurate mental models, and self-confirming equilibrium: a new understanding of the American Revolution. *Journal of Theoretical Politics* **18**, 384–415.
- Friedell M.F.** 1967. On the structure of shared awareness. Working paper.
- Friedell M.F.** 1969. On the structure of shared awareness. *Behavioral Science* **14**, 28–39.
- Fudenberg D. and D.K. Levine** 1993. Self-confirming equilibrium. *Econometrica* **61**, 523–545.
- Gilboa I. and D. Schmeidler** 1995. Case-based decision theory. *Quarterly Journal of Economics* **110**, 605–639.
- Gintis H.** 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, NJ: Princeton University Press.
- Gintis H.** 2016. *Individuality and Entanglement: The Moral and Material Bases of Social Life*. Princeton, NJ: Princeton University Press.
- Gold N.** 2012. Team reasoning, framing and cooperation. In *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour*, ed. S. Okasha and K. Binmore, 185–212. Cambridge: Cambridge University Press.
- Gordon R.M.** 1986. Folk psychology as simulation. *Mind and Language* **1**, 158–171.
- Greenberg J., S. Gupta and X. Luo** 2009. Mutually acceptable courses of action. *Economic Theory* **40**, 91–112.

- Greene J.P.** 1986. *Peripheries and Center: Constitutional Development in the Extended Polities of the British Empire and the United States, 1607–1788*. Athens, GA: University of Georgia Press.
- Greene J.P.** 1991. *Negotiated Authorities: Essays in Colonial Political and Constitutional History*. Charlottesville, VA: University of Virginia Press.
- Hahn F.H.** 1977. Exercises in conjectural equilibria. *Scandinavian Journal of Economics* **79**, 210–226.
- Hayek F.A.** 1937. Economics and knowledge. *Economica* **4**, 33–54.
- Hayek F.A.** 1952. *The Sensory Order: An Inquiry Into the Foundations of Theoretical Psychology*. Chicago: Chicago University Press.
- Hédoin C.** 2014. A framework for community-based salience: common knowledge, common understanding and community membership. *Economics and Philosophy* **30**, 365–395.
- Hédoin C.** 2017. Institutions, rule-following and game theory. *Economics and Philosophy* **33**, 43–72.
- Hindriks F. and F. Guala** 2015. Institutions, rules, and equilibria: a unified theory. *Journal of Institutional Economics* **11**, 459–480.
- Kaneko M. and A. Atsui** 1999. Inductive game theory: discrimination and prejudices. *Journal of Public Economic Theory* **1**, 101–137.
- Kaneko M. and J.J. Kline** 2008. Inductive game theory: a basic scenario. *Journal of Mathematical Economics* **44**, 1332–1363.
- Lewis D.** 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Lewis D.** 1978. Truth in fiction. *American Philosophical Quarterly* **15**, 37–46.
- Matza D.** 1964. *Delinquency and Drift*. New York, NY: John Wiley and Sons.
- Mongin P.** 2016. Spurious unanimity and the Pareto principle. *Economics and Philosophy* **32**, 511–532.
- North D.C.** 2005. What is missing from political economy. In *The Oxford Handbook of Political Economy*, ed. B.R. Weingast and D.A. Wittman, 1003–1010. Oxford: Oxford University Press.
- Perea A.** 2007. A one-person doxastic characterization of Nash strategies. *Synthese* **158**, 251–271.
- Perea A.** 2012. *Epistemic Game Theory: Reasoning and Choice*. Cambridge: Cambridge University Press.
- Perea A.** 2014. From classical to epistemic game theory. *International Game Theory Review* **16**, 1–22.
- Perea A.** 2017. Common belief in rationality in games with unawareness. Working paper.
- Rakove J.N.** 1996. *Original Meanings: Politics and Ideas in the Making of the Constitution*. New York, NY: Knopf.
- Rakove J.N., A.R. Rutten and B.R. Weingast** 2005. Ideas, interests, and credible commitments in the American Revolution. Working paper.
- Reid J.P.** 1976. In accordance with usage: the authority of custom, the Stamp Act debate, and the coming of the American Revolution. *Fordham Law Review* **45**, 335–368.
- Reid J.P.** 1995. *Constitutional History of the American Revolution*. Madison, WI: University of Wisconsin Press.
- Savage L.J.** 1954. *The Foundations of Statistics*. New York, Wiley.
- Schipper, B.C.** 2018. Discovery and equilibrium in games with unawareness. Working paper.
- Sillari G.** 2005. A logical framework for convention. *Synthese* **147**, 379–400.
- Sillari G.** 2008. Common knowledge and convention. *Topoi* **17**, 29–39.
- Sillari G.** 2013. Rule-following as coordination, a game-theoretic approach. *Synthese* **190**, 871–890.
- Simon H.A.** 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* **69**, 99–118.
- Skyrms B.** 2014. *Evolution of the Social Contract*. Second edition. Cambridge: Cambridge University Press.
- Smith A.** 1759. *The Theory of Moral Sentiments*. In *The Glasgow Edition of the Works and Correspondence of Adam Smith*, ed. D.D. Raphael and A.L. Macfie. 1976. Oxford: Clarendon Press.
- Sugden R.** 2011a. Review of H. Gintis's *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. *Economica* **78**, 586–588.
- Sugden R.** 2011b. Salience, inductive reasoning and the emergence of conventions. *Journal of Economic Behavior and Organization* **79**, 35–47.
- Turner M.** 2001. *Cognitive Dimensions of Social Science*. New York, NY: Oxford University Press.
- Vanderschraaf P.** 1998. Knowledge, equilibrium and convention. *Erkenntnis* **49**, 337–369.
- Vanderschraaf, P. and G. Sillari** 2014. Common knowledge. In *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), ed. E.N. Zalta. <https://plato.stanford.edu/archives/spr2014/entries/common-knowledge>.
- Wood G.S.** 1966. *The Creation of the American Republic, 1776–1787*. New York, NY: W.W. Norton.

**Young H.P.** 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

**Zuckert M.** 2005. Natural rights and imperial constitutionalism: the American Revolution and the development of the American amalgam. *Social Philosophy and Policy* **22**, 27–55.

**Nicola Campigotto** is a post-doctoral research fellow at the University of Verona. He holds a PhD in Economics jointly awarded by the Universities of Florence, Pisa and Siena. His current research concerns the evolution of cooperative behaviour and the nature and dynamics of conventions and social norms.