

# Collective intentionality in economics: making Searle's theory of institutional facts relevant for game theory

CYRIL HÉDOIN

REGARDS, University of Reims Champagne-Ardenne

**Abstract:** Economic theories of team reasoning build on the assumption that agents can sometimes behave according to beliefs or preferences attributed to a group or a team. In this paper, I propose a different framework to introduce collective intentionality into game theory. I build on John Searle's account, which makes collective intentionality constitutive of institutional facts. I show that as soon as one accepts that institutions (conventions, social norms, legal rules) are required to solve indetermination problems in a game, it is necessary to assume a form of collective intentionality that comes from what I call *a common understanding of the situation among the players*. This common understanding embodies the epistemic requirements for an institution to be a correlated equilibrium in a game. As a consequence, I question recent claims made by some economists according to which game-theoretic accounts of institutions do not need to assume collective intentionality.

**Keywords:** epistemic game theory, common understanding, collective intentionality, John Searle, institutions

**JEL Classification:** A12, B40, B52, C70

Given that the foundational problem of the multiplicity of equilibria in games has still not been solved, game theorists acknowledge that the extension of Bayesian decision theory to strategic interactions is far from straightforward. While rational choice theory and game theory are still important pieces in the toolkit of most economists, many decision and game theorists are trying to expand the traditional set of assumptions underlying the use of these tools to account for both

---

**AUTHOR'S NOTE:** A previous version of this article was presented at the seminar "Institutionnalismes" at the University of Reims (2012) and at the joint conference AHE, IIPPE, FAPE held in Paris (2012). I particularly thank Nicolas Brisset and Fabien Tarrit for their comments. I also thank Herbert Gintis and another referee for their suggestions and remarks. Finally, I thank François Claveau for his helpful editorial advices. Errors and omissions remain mine.

theoretical and empirical issues. The recent introduction of collective intentionality in a game theoretic framework is an illustrative instance of such a tendency.

Theories of collective intentionality have been developed by philosophers for many decades. Their aim is to provide a convincing explanation for the way individuals coordinate and cooperate by pursuing collective goals and by forming collective intentions or beliefs. In economics, theories of *team reasoning* have been also proposed (Bacharach 2006; Sugden 2000; 2003); they build on the assumption that sometimes agents can behave according to beliefs or preferences attributed to a group or a team. The main motivation for the introduction of team reasoning into economics is related to some of the conceptual difficulties faced by game theory, in particular the problem of the multiplicity of Nash equilibria. A player who team-reasons is seeing the game under a different frame; the decision problem he has to solve is not an individual one, but a collective one. This framing process consists for each agent in transforming the level of agency, i.e., to act according to preferences that are not ascribed to an individual (himself) but to a collective (the team). In some cases, team reasoning allows players to coordinate on one determinate equilibrium.

A distinguishing feature of these theories is that they define ‘I-mode’ and ‘we-mode’ of reasoning as substitutes. Though team-reasoning theories are highly innovative and important, I see at least two limits with this last feature: firstly, team-reasoning does not always suppress the problem of the multiplicity of equilibria. For instance, in pure coordination games, using team-reasoning will not help players to coordinate. Secondly, in the case where the interests of players are not perfectly aligned (i.e., in mixed-motives games), it is not clear why players would or should team-reason. More fundamentally, in such cases, players must agree on a common ordering of collective preferences and, until now, very few contributions have tackled this difficult issue.<sup>1</sup>

In this article, I propose a different framework to introduce collective intentionality into game theory. I build on John Searle’s theory of institutional facts, which makes collective intentionality constitutive of institutional facts. I show that as soon as one accepts that institutions (conventions, social norms, legal rules) are required to solve indeterminacy problems in a game, it is necessary to assume a form

---

<sup>1</sup> Two significant exceptions are Sugden 2000; and Hakli, et al. 2010.

of collective intentionality that comes from what I call a *common understanding of the situation among the players*. This concept embodies the epistemic requirements for an institution to be a correlated equilibrium in a game and can be related to Searle's collective intentionality.

This approach is valuable in at least two respects. Firstly, even though team reasoning may explain cooperation and coordination in some circumstances, institutions are arguably a more pervasive and general explanation. Secondly, by introducing collective intentionality in game theory, I connect Searle's theory of institutional facts with economic theory. Despite the fact that Searle's theory tackles important issues regarding the nature of institutions, his approach has generally been either misunderstood or ignored by economists. A recent article by J. P. Smit, Filip Buekens, and Stan du Plessis (2011) illustrates the standard rational choice objection made against Searle's concept of collective intentionality. My article offers some counterarguments to this critique.

The paper is divided into five sections. The first section presents John Searle's account of collective intentionality and notes that it encompasses all cases where individuals behave according to an institution. In the next section, I interpret institutions as correlated equilibria in a game. I make explicit the epistemic requirements needed for institutions to signal to individuals what they ought to do. The third section shows that these epistemic requirements are fulfilled when players have a common understanding of the situation. I suggest that common understanding has the same properties than Searle's collective intentionality. I build on this point to discuss, in the fourth section, the recent critique of Searle's account by Smit, Buekens, and du Plessis (2011) who argue that an economic explanation of institutions does not need a concept of collective intentionality. A final section briefly concludes.

### **JOHN SEARLE'S ACCOUNT OF COLLECTIVE INTENTIONALITY**

Searle has developed his theory of collective intentionality in several writings (see, in particular, Searle 1990; 1995; 2010). This theory is located at the crossroads of Searle's theory of mind and social theory and a full explanation of it would require us to go deeply into the complexities of his accounts of both. Since this is not possible here, I will confine myself to the elements that are the most important when

one attempts to introduce collective intentionality into a game-theoretic framework. In particular, psychological and cognitive features are secondary, while the link between collective intentionality and social facts is of the utmost importance. Still, an essential building block of Searle's account is its naturalistic stance, i.e., that any assumption or explanation regarding linguistic or social phenomena must be consistent with our current knowledge of natural science, in particular physics and biology (Searle 1995, 5-7). As a consequence, a crucial assumption for any acceptable theory of collective intentionality is that any kind of intentional state must be located in individual minds. Moreover, an intentional state (such as a desire, an intention or a belief) must be taken as independent of the actual state of the world; an individual may hold radically wrong beliefs or build his intention on a mistaken conception of the world (Searle 1990). As Searle himself notes, these commonsensical and pre-theoretical requirements are fully in line with all the variants of methodological individualism.

Given these constraints, Searle argues that, in addition to singular or individual intentionality, the ontology of our world is also made of another form of intentionality: collective intentionality. According to Searle, collective intentionality is a *primitive* biological phenomenon that is not an exclusive property of humans but that is also possessed by other animals (Searle 1995, 24). In Searle's social theory, collective intentionality is deemed to be constitutive of any social fact (Searle 1995, 26).

The main implication is that, according to Searle, one cannot reduce collective intentionality to some form of singular intentionality. In particular, Searle (1990) rejects popular accounts of 'we-intentions' that reduce the proposition 'we intend that' to a set of 'I intend that' augmented by a set of mutual beliefs regarding the intentions of everyone, i.e., accounts that equal 'we intend that x' with the infinite hierarchies of propositions 'I intend that x given that I believe that you intend that x, you intend that x given that you believe that I intend that x, I believe that you believe that I intend that x, you believe that I believe that you intend that x', and so on *ad infinitum*, effectively making the intention that x common knowledge in the population.<sup>2</sup>

---

<sup>2</sup> Such a reductive and individualistic account of collective intentionality has been particularly developed by Tuomela and Miller 1988. Note that despite the fact that I use the concept of common knowledge below, my own account is very different from the one of Tuomela and Miller. However, as one of the referees remarked, I shall note that Raimo Tuomela's most recent writings on the topic of collective intentionality

Searle gives several arguments against this kind of reduction. The most decisive one is that while it tries to make collective intentionality the mere aggregation of individual intentional states, it fails to account for the fact that individual intentionality is *derived* from the collective intentionality that members of a group may share. Indeed, in many cases, I intend that  $x$  *because* we intend that  $x$  (or some other propositional content  $x'$ ).

To understand the structure of Searle's framework, it is useful to start with the way Searle defines intentionality. Intentionality is "the capacity of the mind to represent objects and states of affairs in the world other than itself" (Searle 1995, 6-7; see also Searle 2010, 25). According to this definition, intentions, desires, and beliefs are all kinds of intentional states. For example, desires are an instance of a mind-to-world relationship where one wants something to become actually the case. Beliefs are rather an instance of world-to-mind relationship, since a belief is about something that is already the case. In both cases, the intentional state is about an external object.

Though Searle's analysis essentially focuses on a specific form of intentional states, namely intentions-in-action (i.e., intentions that cause an action in a self-referential way), it is perfectly reasonable to extend it to other intentional states. Still, making the simplifying assumption that intentions always lead to the corresponding actions, discussing the case of intentions-in-action is sufficient for our purpose. Searle's analysis of the various forms of intentionality uses the simple notational device  $S(p)$ , where  $S$  stands for the psychological or intentional state (an intention-in-action, a belief, a desire) and  $p$  denotes the propositional content and consequently determines the conditions of satisfaction of the state. In the simple case of an action where one is raising one's arm, we have the following expression, where **ia** stands for *intention-in-action*:

(1) **ia** (this **ia** causes: my arm goes up) CAUSES: MY ARM GOES UP

The first term outside the parentheses denotes the intentional state. The expression inside the parentheses depicts the "mental" component of the action. It states the conditions of satisfaction of the intentional state which, since we are concerned here about an intention, is self-referential: one of the conditions of satisfaction of the state is that the

---

(e.g., Tuomela 2007) largely give up the individualistic stance and are much closer to the account developed here. Tuomela has also written extensively on how Searle's account and his own are related, e.g., Tuomela 2011.

state itself causes an event of the type represented in the rest of the conditions of satisfaction. The expression in capital letters is a representation of the “physical” component of the action. For the remainder of this section, I will ignore the physical component and assume that it always satisfies the content of the mental component.

The same expression may be used to represent an intention causing an action that causes itself a further action, i.e., a by-means-of relation. It can also represent an intention causing an action which is constitutive of a further action, i.e., a by-way-of relation. As an example of the former, in the case where one intends to fire a gun by pulling the trigger, we have:

(2) **ia** B by means of A (this **ia** causes: A [trigger pulls], causes: B [gun fires])

This reads as ‘I have an intention-in-action to cause B (fire the gun) by means of A (pulling the trigger), this intention-in-action leads me to have the intention to pull the trigger to make the gun fire’.

The point is now to see how this notation and the theoretical framework on which it is grounded can extend to cases of collective intentionality. Here enters Searle’s main contention: while individuals are regularly in a singular-intentional state such as ‘I intend’, ‘I desire’ or ‘I believe’, we also routinely use first-person plural forms of intentionality such as ‘we intend’, ‘we desire’ and ‘we believe’. As noted, Searle takes this form of collective intentionality to be irreducible to singular forms and, moreover, to be *causally* responsible for the formation of individual intentional states (i.e., ‘I intend that... *because* we intend that...’).

Take the example of a play in an American football game where an offensive lineman intends to block defensive players. The lineman’s intention is caused by the fact that he participates in a football game and that it has been decided that the offense he is part of will intend to complete a pass to gain some yards. Originally, Searle (1990) proposed the following general formulation in terms of an achieve-collective-B-by-means-of-singular-A type:

(3) **ia** collective B by means of singular A (this **ia** causes: A [block defensive players], causes B [completion of the pass play])

Here again, the expression outside the parentheses denotes the type of intentional state. Therefore, the collective dimension of the state

features explicitly. It reads as ‘we intend B by means of A’, where B is the collective target of the intention, while A is the individual means. Since for Searle, intentionality is the result of a brain process, this literally means that the collective intentionality is “in the head” of each individual. However, this original account of collective intentionality faces a great difficulty.

Indeed, it should be clear that in the formulation above A can only cause B under certain circumstances, in particular if others have the appropriate intentional states. Therefore, the propositional content inside the parentheses is unacceptable (Bardsley 2006, 137). This difficulty comes from the fact that any account of collective intentionality must satisfy what is generally called in the literature the *uncontroversial constraint*: an individual’s intentions (or any other intentional states) cannot be said to range over others’ actions. As a consequence, it seems impossible to have a collective intention (‘we intend’) or a collective belief (‘we believe’), because one’s intentional state cannot cause others’ actions. Stated in this way, Searle’s account—as well as all other non-reductive accounts of collective intentionality—faces a dilemma: either violating the uncontroversial constraint or giving up the non-reductionist stance (see Bardsley 2006). Both alternatives are hardly desirable, since the former would imply a barely sustainable metaphysical assumption, while the latter would be in contradiction with the essence of a non-reductive account of collective intentionality.

In a full chapter devoted to his theory of collective intentionality, Searle (2010, 42-60) reformulates his account to make clear that the tacit presupposition that everyone will be doing his part in the collective endeavour is not part of the propositional content but rather is an implicit statement constitutive of the collective intentionality. Searle’s new account is thus the following (2010, 53):

- (4) **Bel** (my partners in the collective also have intentions-in-action of the form (**ia** collective B by means of singular A (this **ia** causes: A [block the defensive players], causes: B [completion of the pass play])))

We have thus added an extra clause **Bel** that reads as follows: ‘I have a belief that my partners in the collective also have the same collective intention-in-action as mine’. Searle underlines that this belief is not a part of the *content* of the intention-in-action, thus satisfying the uncontroversial constraint. Rather, it is an integral part of the collective

intention: to have a collective intention implies to believe that others have the same collective intention. Incidentally, reading Searle literally, this belief seems *individual*: it is about others' intentional states but it is not a 'we believe that' type.

I do not regard this new account as fully satisfactory. First, we-intentionality can no longer be considered non-reducible because individual beliefs are necessary. Second, and more importantly, the extra clause **Bel** seems insufficient: even if I believe that you have the same collective intention than me, I may doubt that you have the same belief about me. In this case, I may refrain to form an individual intention on the basis of this collective intention. By this reasoning, we reach the inescapable conclusion that for a collective intention to be effective, we must have an infinite set of iterated mutual beliefs about our intention, ultimately grounding the collective intention on a singular form of common belief. As I suggest in the next two sections, there is one escape though if we interpret Searle's **Bel** condition not in the traditional epistemic sense but as a dispositional and pre-intentional property. In this case, the very possibility of collective intentionality is a feature of what Searle calls the "Background". I further discuss this point below.<sup>3</sup>

Crucially, collective intentionality (in particular collective intentions) is constitutive of institutional facts.<sup>4</sup> According to Searle's taxonomy of facts (Searle 1995, 121), social facts involve a form of cooperation which requires collective intentionality. Since institutional facts are a subclass of social facts where status functions are assigned to brute phenomena (i.e., 'this piece of paper counts as a twenty dollar bill'), they also rely on a non-reductive form of collective intentionality. In particular, Searle argues that the maintenance of institutions through time necessitates

---

<sup>3</sup> Searle's new account also features a more explicit discussion of what he calls "collective recognition" or "collective acceptance". He defines collective recognition as a weaker form of collective intentionality which operates even when social agents do not have intentions to cooperate. Collective recognition is an important category because it underlies all of Searle's theory of institutional facts, in particular his concepts of status functions and of constitutive rules. Unfortunately, this contributes to muddle the water somewhat, in part because it fits uneasily with Searle's own framework developed in *The construction of social reality* (1995). In this book, Searle explicitly indicates that cooperation is constitutive of all social facts, of which institutional facts are a subset. Since Searle's definition of cooperation is fairly large (see Searle 1995, 23-24), it is difficult to point to (non-social) facts involving collective recognition but not collective intentionality.

<sup>4</sup> Note that Searle's theory of institutional facts also makes the strong claim that language is also constitutive of institutional facts. For a function to be assigned, it has to be represented by some form or another of symbolism. I will not put this idea under scrutiny here but I will have to go back to it in the last section.



collective intentionality (Searle 2010, 57). In the next two sections, I will show that a similar non-reducible account of collective intentionality is also relevant in a rational choice framework.

### **INSTITUTIONS AS CORRELATED EQUILIBRIA: MUTUAL AWARENESS AND SYMMETRIC REASONING**

Building on Searle's account of collective intentionality, this section and the next demonstrate that it is fruitful to assume collective intentionality in rational-choice and game-theoretic frameworks. In fact, it is even necessary if one accepts the necessity of institutions to account for coordination and more generally for agents' choices in strategic interactions.

Our socioeconomic life is full of situations mixing coordination and conflict. Even if pure coordination problems are also pervasive, social interactions where agents have to coordinate while having conflicting interests are particularly difficult and important. The difficulty is that in such interactions, efficiency makes coordination desirable but equity concerns and conflicting preferences over outcomes may make coordination particularly hard to achieve.

Social scientists, including economists, tend to agree that *institutions* are one of the main devices used by agents in mixed-motive games to form beliefs and to act. Institutions are usefully defined as *social norms, conventions, legal rules or formal organizations that generate a set of consistent beliefs in a population such that a stable behavioural pattern is observable through time*. In other words, institutions are kinds of social objects that enter as an input in the agents' intentional states to help them form consistent beliefs regarding the state of the world and regarding what others will do.

In the remaining of the paper, I will follow Herbert Gintis (2009) in formalizing institutions as *correlated equilibria* in games. The main motivation for using this solution concept rather than the more classical Nash equilibrium solution is twofold. Firstly, as Aumann and Brandenburger (1995) have emphasized, the epistemic conditions that must be satisfied for rational players to play a Nash equilibrium are very demanding, especially with three or more players. Indeed, both common knowledge of expectations and common priors over the way the game will be played must be satisfied. Except under highly specific circumstances, these conditions are unlikely to be satisfied. Secondly, and more fundamentally, norms and other kinds of

institutions are more properly interpreted as implementing correlated equilibria. An important theorem by Robert Aumann (1987) shows that the strategy profile played by Bayesian rational players with common priors forms a correlated equilibrium. If one accepts the assumption that common priors are a necessary condition for an institution to exist, then formally institutions are correlated equilibria.<sup>5</sup>

My main claim is that for an institution to serve as a correlating device in a game, we have to assume a form of collective intentionality qualitatively similar to Searle's. I will work through the following mixed-motive game, which I will call the 'property game' (see Table 1).

**Table 1: The property game**

		Bob	
		attack	negotiate
Ana	attack	$(V-C)/2; (V-C)/2$	$V; 0$
	negotiate	$0; V$	$V/2; V/2$

The property game (also known as the chicken game) depicts an interaction where two agents are fighting for an asset (a prize, a territory) of value  $V$ . Each player has two strategies available: either to attack in order to take the asset by force or to negotiate peacefully. When one of the players attacks while the other attempts to negotiate, the former wins the prize for sure. If both negotiate, they share the asset equally. Finally, if both attack, each player wins with probability  $\frac{1}{2}$  and the loser suffer a loss of  $-C$  which might be interpreted as an injury. We assume that  $C > V$ .

As usual, this game structure is common knowledge among the players, the players are rational and this is common knowledge. This game embodies an element of coordination that is captured by the fact that it has two Nash equilibria in pure strategies: [attack; negotiate] and [negotiate; attack]. There is also a third, mixed-strategy Nash equilibrium where each player attacks with probability  $V/C$ . Because  $C > V$ , each player prefers losing the asset rather than fighting when the

---

<sup>5</sup> Note that using the concept of correlated equilibrium does not entail the rejection of the Nash equilibrium solution, nor does it imply making non-standard assumptions regarding rationality or knowledge in a game. Indeed, every correlated equilibrium corresponds to a Nash equilibrium in the larger game where Nature moves first to signal the state of the world to the players.

other attacks. However, the players' preferences over the outcomes are clearly inconsistent since each prefers to take the asset over giving it up.

The difficulty lies in the fact that it is common knowledge among the players that their preferences over the outcomes are inconsistent. In such a case, an institution (a social norm or a legal rule) is required to settle the dispute. Typically, an institution will select one of the equilibria; however, in our case, there is no reason why an institution should adjudicate the conflict between Ana and Bob in favour of the former or of the latter. Indeed, who will be *entitled* to claim the asset depends on the specific situation of the two players. For instance, the norm might distinguish between the *actual* possessor of the asset (the incumbent) and the non-possessor (the challenger). Assume that the distinction between incumbent and challenger is non ambiguous and common knowledge and assume that each player is equally likely to be the incumbent or the challenger; we can then provide the alternative description of the game represented in Table 2.

**Table 2: The property game with incumbent and challenger**

		Bob			
		always attack	always negotiate	attack if incumbent	attack if challenger
Ana	always attack	$(V-C)/2$	$V$	$(3V-C)/4$	$(3V-C)/4$
	always negotiate	$0$	$V/2$	$V/4$	$V/4$
	attack if incumbent	$(V-C)/4$	$3V/4$	$V/2$	$(V-C)/2$
	attack if challenger	$(V-C)/4$	$3V/4$	$(V-C)/2$	$V/2$
<i>N.B. Only the payoffs of the row player are shown.</i>					

Once the players are able to distinguish between being the incumbent and being the challenger, they can use *conditional* strategies, i.e., strategies that associate an action to a state of the world. Framing the 'property game' this way does not solve the problem of the multiplicity of equilibria since there are still two equilibria in pure strategy<sup>6</sup> and one equilibrium in mixed-strategy. However, now that

<sup>6</sup> These are: [attack if incumbent; attack if incumbent] and [attack if challenger; attack if challenger].

the distinction between incumbent and challenger has been made, an institution that settles the dispute along an intuitively more satisfactory impersonal criterion can be defined. In particular, a *property norm* will select the [attack if incumbent; attack if incumbent] equilibrium, avoiding costly conflicts that arise with probability  $(V/C)^2$  when the players play the mixed-strategy equilibrium.<sup>7</sup>

I will now examine how the property norm effectively changes the players' beliefs. To acknowledge the role played by the features regarding the knowledge and the beliefs of the players, it is useful here to define the *epistemic* property game. An epistemic game retains the classical feature of a game (i.e., a set of players, a set of pure strategies and a consequence function), but adds an explicit mapping of the players' knowledge about the states of the world (see Gintis 2009, 83-84).

Define  $\Omega$  as the set of possible states of the world and  $P_i$  the possibility sets of  $\Omega$  for each player  $i$ . Possibility sets partition the set of worlds  $\Omega$  into units of knowledge defined by a knowledge partition. A possibility set indicates which states of the world  $\omega' \in \Omega$  a player thinks is possible when the actual state of the world is  $\omega \in \Omega$ , which we denote as  $P_i\omega$ . Finally, each player is endowed with a subjective prior  $p_i(\cdot; \omega)$  that is a function of the state of the world  $\omega$ . A player's subjective prior defines, among other things, his belief over the strategy profiles to be played by the other players at  $\omega$ . A Bayesian rational player will choose the strategy that maximizes his expected payoff given his conjectures regarding the state of the world and how others will play. Finally, we identify an *event*  $E$  as the set of possible worlds in which that event happens. An agent *knows* an event  $E$  when every world the agent considers possible is a subset of  $E$ . The event of an agent  $i$  knowing that  $E$  is denoted  $K_iE$ .<sup>8</sup> Formally,  $K_iE = \{\omega \mid P_i\omega \subseteq E\}$ .

Now apply the above formalism to the 'property game'. We distinguish between two states of the world  $\Omega = \{\omega_1; \omega_2\}$  which obtain with an equal probability  $q_\omega = 1/2$ . Assume that  $\omega_1$  is the state of the

---

<sup>7</sup> There is an evolutionary argument for the selection of one of the equilibria with conditional strategies beyond the mere intuitive plausibility. Indeed, in an evolutionary setting, the mixed-strategy is not evolutionary stable. Quite the contrary, both conditional strategies are evolutionary stable. An intriguing fact however is that the anti-property strategy 'attack if challenger' has an even chance to spread in the population than the more intuitive property strategy.

<sup>8</sup>  $K$  is a knowledge operator. As usual in epistemic game theory, I assume that the knowledge operator satisfies the requirements of the modal logic S-5. See Binmore 2007, chapter 12, for an accessible presentation.

world where Ana is the incumbent while she is the challenger in  $\omega_2$ . My assumption that the distinction incumbent/challenger is unambiguous is equivalent to the formal statement that  $\omega_1$  and  $\omega_2$  are in different cells in the knowledge partition of both players, i.e., when the actual state is  $\omega$ , each player knows this for sure. Therefore, the events  $E = \{\omega_1\}$  and  $E = \{\omega_2\}$  are *public events*. When E happens, it is self-evident for each player that E happens, i.e.,  $E = K_i E$  for all  $i$ . Since  $\omega_1$  and  $\omega_2$  are also in two different cells in the communal possibility sets of all  $i$ ,<sup>9</sup> E is common knowledge because everyone knows that E is self-evident for everyone:  $i$  knows that E obtains,  $i$  knows  $j$  knows that E obtains,  $i$  knows that  $j$  knows that  $i$  knows that E obtains, and so on. Formally,  $E \subseteq CK(E)$ ; for all E and with  $CK(E)$  the event that (Everyone knows that)<sup>∞</sup> E. The property norm described above may now be restated as follows:

$$N(E) = (\text{if } \omega_1, [\text{attack}; \text{negotiate}], \text{if } \omega_2, [\text{negotiate}; \text{attack}])$$

Our notation  $N(E)$  indicates that event E is governed by norm N (see Gintis 2009). Generally speaking, a norm  $N(E) = (s_1, \dots, s_n)$  specifies a vector of strategies for the  $n$  players such that no Bayesian rational player can gain by playing another strategy  $s_i'$  when E obtains. The norm defines a *correlated equilibrium* because the strategies are correlated to a specific state of the world. Crucially, at equilibrium, the norm  $N(E)$  itself is common knowledge among the players, thus  $CK(E) \subseteq CK(N)$ .

This expression depicts an *indication relation*. To say that an event E is governed by a norm N is the same as to say that E *indicates* N to the players. The notion of indication relation is central in David Lewis's theory of common knowledge (Lewis 2002 [1969]; Cubitt and Sugden 2003) and captures the fact that individuals have to be able to infer a norm from a particular event for that norm to be effective.

This framework makes explicit the epistemic properties of an institution. The players' subjective priors lead them to form a conjecture  $\phi_i^\omega$  regarding how the other players will play at  $\omega$ . Thus, we can define a belief profile  $\phi(\omega) = (\phi_1^\omega, \dots, \phi_n^\omega)$  correlated with the state  $\omega$ , such that Bayesian rational players will respond by a strategy profile  $s(\omega) = (s_1^\omega, \dots, s_n^\omega)$  where everyone is maximizing expected utility. In the 'property game', the property norm leads everyone to conjecture

---

<sup>9</sup> The communal possibility sets are defined as the *meet* of the possibility sets of each player. Obviously, since the possibility sets of all players distinguish  $\omega_1$  from  $\omega_2$ , the communal possibility sets have also this property.

that the incumbent will attack: it correlates the state of the world to a set of conjectures and of strategies, i.e.,  $N(E) = \phi(\omega)$ . Thus, we have a formal expression of the ability of an institution to generate a set of beliefs in a population.

However, this is not the end of the story. We have assumed the indication relation  $CK(E) \subseteq CK(N)$  while in fact it embodies crucial epistemic assumptions. Specifically, at least two additional assumptions that go beyond the classical framework of Bayesian decision theory are needed (Gintis 2009): *mutual awareness* and *symmetric reasoning*. I argue that both can be accounted for as properties of a form of collective intentionality such as envisioned by Searle.

The first thing to note is that, in my example, all events are public events. That means that when E obtains, each player knows for sure that E obtains. But to be allowed to infer from the mutual knowledge of E that E is common knowledge, we have to make the seemingly benign assumption that each player knows the knowledge partition of the other player. This assumption is implicit in the above model because a description of each state of the world includes, among other things, a description of the agents' possibility sets. However, this mutual awareness assumption may be really strong in some settings. In my example, it consists in hypothesizing that the incumbent/challenger distinction is *really* unambiguous and that Ana and Bob both know that they exactly define in the same way both concepts. Moreover, it means that if Ana knows she is the incumbent, she must be able to infer with a total confidence that it is unambiguous for Bob that he is the challenger.

Regarding symmetric reasoning, it is crucial to note that the ability for a norm  $N(E)$  to become common knowledge when an event E is publicly observed relies on the fact that everyone infers N from E. We say that two persons  $i$  and  $j$  are symmetric reasoners if and only if  $i$  can infer from his knowledge that  $x$  and that  $j$  knows that  $x$ , that  $j$  also knows  $y$ . More formally, for all  $i, j$  and for any events  $E, N \subseteq \Omega$ ,

$$(K_i E \subseteq K_i N) \wedge (K_i [K_j E]) \Rightarrow K_i [K_j N].$$

Such an assumption is at the heart of the indication relation pioneered by Lewis, who supposes that individuals share a set of modes of reasoning (Cubitt and Sugden 2003; Lewis 2002 [1969]; see also Gintis 2009 and Vanderschraaf 1998). Moreover, considering that individuals can deliberate over the modes of reasoning used by members of a

population, we have a second-order problem of coordination where each person is uncertain regarding the way others reason. The problem of course is that there is no reason to suppose that there is only one way to infer  $y$  from  $x$ , even if deduction is available. People routinely use other modes of inference as a function of the situation in which they interact.

One way to circumvent the problem of mutual awareness and symmetric reasoning would be to assume that the players in an epistemic game have *common priors*. The common prior assumption (also known as the Harsanyi doctrine) simply consists in taking for granted that all players  $i$  in a population share the same subjective prior  $p_i(\cdot; \omega)$ . It is then a theorem that Bayesian rational players with common priors will implement a correlated equilibrium in a game where Nature chooses the state of the world (Aumann 1987). As noted above, this theorem is the main motivation for using the concept of correlated equilibrium. But the common prior assumption is clearly disputable, at least when we do not restrict ourselves to phenomena where objective probabilities can be inferred from law-like regularities.<sup>10</sup>

In particular, combined with the assumption that individuals' posterior beliefs are common knowledge, the Harsanyi doctrine leads to the claim that it is impossible to “agree to disagree” (Aumann 1976). This claim clearly does not stand up against empirical evidence and must be considered more a formal curiosity than a substantive insight. Still, it remains true that people are largely able to coordinate in their daily activities and even on more fundamental problems such as agreeing on principles of justice (e.g., Binmore 1998). What I want to suggest is that something akin to the Searlian collective intentionality might provide a satisfactory explanation of mutual awareness and symmetric reasoning.

### COMMON UNDERSTANDING AND COLLECTIVE INTENTIONALITY IN GAMES

Since David Lewis's study of conventions (Lewis 2002 [1969]), it is standard to assume that norms and more generally institutions are devices that select a Nash equilibrium in a game. However, I follow Gintis (2009) in considering that it may be more relevant to define

---

<sup>10</sup> This is the essence of Binmore's (2009) critique of Bayesianism. Following the seminal work of Leonard Savage, Binmore contends that Bayesian decision theory only operates on secure grounds when applied to small worlds, *i.e.* worlds where we know that we cannot be surprised. In large worlds, assuming common priors and asserting the absolute validity of Bayesian updating of subjective probabilities is foolish.

institutions as correlated equilibria. My main motivation in doing so is methodological: to define an institution as a correlated equilibrium leads one to reflect on the epistemic properties of the institution. As the preceding section showed, this approach reveals non trivial epistemic assumptions routinely made in game theory, in particular mutual awareness and symmetric reasoning. I now suggest that one can interpret mutual awareness and symmetric reasoning as peculiar instances of Searle's collective intentionality.

Hédoin (2012) proposes to frame the assumption of symmetric reasoning as referring to cases where agents have a *common understanding of the situation*. Basically, the point is that to infer a set of conclusions  $C$  from a situation defined by an event  $E$ , the members of the population will use a set of modes of reasoning and of background information  $B$ . If every member of a population infers  $C$  from  $E$  on the basis of  $B$ , then we say that members of the population have a *shared-reflexivity* of the situation. More formally,

*( $\Phi$  -  $P$ )-shared reflexivity*: there is ( $\Phi$  -  $P$ )-shared reflexivity among the members of a population constituted by the set of agents  $P$  if the agents share the same reasoning modes and background information  $B$  with respect to a non-empty set of events  $\Phi$ . In practice, all agents in  $N$  infer the same conclusion  $C$  when they observe an event  $E \subseteq \Phi$ .

If for each person  $C$  contains propositions about the conclusions reached by the other members of the population, then  $C$  depends on the set  $B_i$  used by each member  $i$  of the population. We define a *second-order shared-reflexivity of the situation* in the population as the fact that each person  $j$  embeds in  $B_j$  information regarding the sets  $B_i$  of every other person  $i$ . In principle, this leads to an infinite degree of order of shared reflexivity. There is *common understanding of a situation* among the members of a population when the degree of order of shared reflexivity reaches an arbitrarily high level. Formally,

*( $\Phi$  -  $P$ )-common understanding*: there is ( $\Phi$  -  $P$ )-common understanding among the members of a population constituted by the set of agents  $P$  if there is an infinite order of ( $\Phi$  -  $P$ )-shared reflexivity. In practice, all agents in  $P$  infer the same conclusion  $C$  when they observe an event  $E \subseteq \Phi$  and this is common knowledge.



When common understanding is obtained in a population, members of this population frame the situation identically and crucially take this fact for granted. Common understanding is a necessary condition for something to become common knowledge in a population. Common understanding is a cognitive disposition that allows the formation of intentional states such as beliefs or intentions. In this sense, it is partly *pre-intentional*. I suggest that the concept of common understanding can be extended to mutual awareness. Indeed, mutual awareness corresponds to proposition:  $K_i E \subseteq K_i [K_j E]$ .

This proposition says that if  $i$  knows  $E$ , then he also knows that  $j$  knows  $E$ . The proposition is grounded on an inference that what is self-evident to  $i$  is also self-evident to  $j$ . Therefore, mutual awareness is no more than a special case of symmetric reasoning defined by an indication relation linking two separated events.

Where common understanding comes from is surely an important question. It is quite probable that it is partly innate. As members of the human species, we share a common evolutionary history and consequently basic cognitive functions and mechanisms. As an example, humans are universally disposed to identify recurrent patterns in nature. However, our propensity to interpret in similar ways public events (and our propensity to be mutually aware of such events) surely rely on the fact that the members of a population will generally share a system of symbolic cues and meaning (Chwe 2003). Hence, common understanding is largely cultural. In particular, common understanding will generally obtain among the members of the same *community*, i.e., a group of persons who interact through the same institutions. By interacting through the same institutions, individuals progressively develop cognitive and behavioural dispositions constitutive of their modes of reasoning.

The apparent circularity of my argument, where common understanding allows the working of institutions and institutions foster common understanding is not a problem. Indeed, the pre-intentional content of common understanding refers to what Searle calls the “Background”, which he defines as “the set of non-intentional or pre-intentional capacities that enable intentional states of function” and where capacities are “abilities, dispositions, tendencies, and *causal structures generally*” (Searle 1995, 129, emphasis in the original).

Searle argues that in the first place the Background enables linguistic interpretation: the set of truths conditions regarding the meaning of

a sentence is always determined given certain Background capacities. But the pre-intentional structure of the Background is also essential in enabling what I propose to call “institutional interpretation”: the social world is full of institutional signs (conventions, rules, norms, signals) that must be interpreted. In the ‘property game’, the content of the property norm has to be interpreted by Ana and Bob. As we have seen, game theory makes this point obscure with the hidden assumptions of mutual awareness and symmetric reasoning. The point is that an institutional sign is always incomplete and must be *interpreted*. Interpretation always takes place on the basis of a more global set of institutions that defines a community. Searle’s Background is captured by the pre-intentional content of the common understanding concept.

Undoubtedly, common understanding also has a conscious and intentional content. To illustrate this claim (which could be easily tested experimentally), a simple appeal to one’s occasional experience of interacting with people endowed with a different culture should be sufficient. When I drive from my home to my University office, I generally follow the institutional signs on the road (i.e., the traffic rules) without much deliberation. I believe that other drivers will stop at red lights and drive on the right because I *take it for granted* that they are acquainted with such institutional signs. In other words, because we are members of my community, I do not have to reflect profoundly on my interpretation of the traffic rules. Now, things are clearly different when I drive my car in other countries. Here, even though institutional signs might be in principle identical (such as red lights), I will (consciously) be much more thoughtful in their interpretation. Basically, since I *know* that I interact with “strangers”, I cannot fully assume common understanding and I will form beliefs regarding their interpretation of institutional signs. Here, I want to suggest that this type of intentional content that is constitutive of common understanding is the product of a collective intentionality of the type conceptualized by Searle.

When a common understanding of a situation is obtained, each participant in the interaction takes for granted that everyone else is interpreting the situation like him.<sup>11</sup> This pre-intentional commitment

---

<sup>11</sup> It is possible to express the idea of common understanding in purely individualistic terms, for example by assuming that symmetric reasoning is common knowledge. In the standard epistemic game-theoretic approach used in the preceding section, each possible state of the world contains everything that is relevant for the players, including possibly the kinds of reasoning. Therefore, once we have partitioned  $\Omega$  into

sets the ground for a *collective* intentional state which can be formulated with the following proposition:

- (5) *Collective intentionality in a game*: as members of population  $P$ , in the situation defined by the game  $G$  and a specific event  $E$ , we intend to play the strategy profile  $s^*$ , therefore  $I$  should intend to play the strategy  $s_i^*$ .

This statement combines two kinds of reasoning: a we-reasoning where each person, as a member of  $P$ , ascribes a collective intentional state to the group of participants in  $G$ ; and an I-reasoning where each person reflects on what he should do given his personal preferences and his collective intentional state. The locution ‘therefore’ in this statement embodies a causal relationship where one does  $s_i^*$ , because everyone is collectively intending  $s^*$  when  $E$  holds in  $G$ . In the ‘property game’, observing event  $E$ , Ana will do  $s_i^*$  (attack or negotiate) *because* it maximizes her expected utility given the fact that she takes for granted that Bob and herself collectively intend  $N(E)$  in  $E$ .

This is in line with Searle’s account of collective intentionality where collective intentional states are causally responsible for individual intentional states. We can make this clearer by reformulating Ana’s and Bob’s intentions to attack or to negotiate in Searle’s notation. The ‘property game’ embodies the kind of general cooperation that is constitutive of social facts according to Searle. What we have here is the following:<sup>12</sup>

- (6) **ia** collective B by way of A (this **ia** causes: A [playing my part of the strategy profile defined by the property norm], constitutes B [cooperation through the property norm])

Since the *content* of the property norm is in the propositional content of (6), both players take it as given. Now, if we add the **Bel** clause, we obtain:

---

several states of the world and defined the players’ information partitions, all we need to solve the game is already there. As we have already indicated, the same result occurs if we assume common priors. The problem is that there is no reason to assume that the set  $\Omega$  can be partitioned into well-defined possible worlds or that, at each possible world  $\omega$ , everyone knows everyone’s possibility sets or modes of reasoning *prior* to any form of interaction or collective recognition.

<sup>12</sup> Since, properly speaking, a set of individual strategies is constitutive of a strategy profile and hence of an institution, the type of collective intention relevant here is not a causal ‘B by means of A’, but rather a constitutive ‘B by way of A’. Searle’s framework allows for the two formulations. This makes no real difference here.

- (7) **Bel** (my partners in the collective also have intentions-in-action of the form (**ia** collective B by way of A (this **ia** causes: A [playing my part on the strategy profile defined by the property norm], constitutes B [cooperation through the property norm])))

Following what has been said in the first section, we must take this belief to be grounded on a pre-intentional disposition. Because the content of this belief includes the collective intention-in-action (6) and since (7) contains the content of the norm, the players actually implement the norm *because* they take for granted that everyone shares the same intention to implement this particular norm with this particular content. Propositions (5) and (7) are two different statements of the same collective intention.

In comparison to theories of team reasoning (Bacharach 2006; Sugden 2000), my account does not need to stipulate team preferences. Indeed, *there is* collective intentionality in the ‘property game’ despite the fact that the matrix only indicates individual payoffs related to individual preferences. Hence, my account avoids one of the most contentious points of the theory of team reasoning, i.e., how team preferences are generated. By contrast, collective intentionality in games helps to explain how the players form their beliefs.

As indicated in the preceding section, an institution  $N(E)$  is formally equivalent to a vector of conjectures correlated to the state of the world  $\phi(\omega) = (\phi_1^\omega, \dots, \phi_n^\omega)$ . As an example again, in the ‘property game’, Ana believes with probability one that Bob will play ‘negotiate’ when  $\omega_1$ ; for the same state of the world, Bob believes with probability one that Ana will play ‘attack’. These *individual* beliefs are generated by the fact that both players have a common understanding of the game and that they share a *collective* intentional state in the situation such that a peculiar property norm applies. Collective intentionality in games is thus a plausible explanation of subjective prior, and in particular of *common* prior.

To end this section, note again that the **Bel** clause in proposition (7) is not a belief in the traditional epistemic sense. Therefore, consistent individual beliefs are formed as a result of the common understanding of the situation which is more likely to obtain when people are ostensibly members of the same community. The pitfall of infinite regress is avoided as soon as one acknowledges that common understanding is ultimately grounded on pre-intentional states.

Incidentally, Searle's earlier writings on collective intentionality linked the ability to form collective intentional states with Background capacities: "it seems to me that the capacity to engage in collective behaviour requires something like a pre-intentional sense of "the other" as an actual or potential agent like oneself in cooperative activities" (Searle 1990, 19). Undoubtedly, it is tempting to postulate a deep causal relation between the 'pre-intentional sense of the other' and the peculiarity of the evolutionary history of the human species which has been shaped by the importance of communities (see Bowles and Gintis 2011). Communities are thus at the same time ultimate and proximate causes in our ability to coordinate and to cooperate.

### **COLLECTIVE INTENTIONALITY AND INSTITUTIONS: SEARLE VERSUS ECONOMIC THEORY?**

In this last section, I will make use of what I have said above to discuss the recent critique made by J. P. Smit, Filip Buekens, and Stan du Plessis (2011) against Searle's theory of institutional facts. Their basic claim is that all the Searlian apparatus, including his account of collective intentionality, is unnecessary to explain and understand institutional objects. As an alternative, the authors propose an approach that is deemed to be more compatible with standard economic theory, and which is grounded on the concepts of actions and incentives. I will not review here the various arguments made by these authors against Searle's theory.<sup>13</sup> I will focus on the authors' alternative approach and claim that while interesting it is not a true alternative to a theory of institutional facts embodying an account of collective intentionality.

The central tenet of the authors' "economic" theory of institutional facts is that institutional facts are tied to a set of actions. Actions are themselves triggered by a set of incentives, i.e., a measure of the relative desirability of the components of a set of possible actions. The authors insist that the origins of the incentives do not matter to characterize the

---

<sup>13</sup> Smit, Buekens, and du Plessis's main critique is precisely directed against Searle's account of collective intentionality. They argue that because Searle gives a role to collective intentionality in the assignment of functions, institutional facts cannot be reduced to natural facts. The authors deny this irreducibility and contend that "the mind-dependence of such objects [institutions and institutional facts] is the result of the need for incentivization, and the fact that talk about incentivization is warranted only if talk about desires and beliefs—and hence minds—is warranted" (Smit, et al. 2011, 4). Note that the authors' claim that Searle's theory is ultimately non-reductionist runs contrary to Searle's claim that his theory is grounded on a naturalistic stance.

nature of the institutional objects. In other words, the source of the incentive may be tied to different causal mechanisms for different persons; what counts is that as a result of the incentive everyone is behaving in a particular way that defines the institutional object. To illustrate their point, the authors start from one of Searle's favourite examples, the emergence and the maintenance of a border. I will myself use Smit and co-authors' treatment of this case to make my own point. Here is Searle's discussion of borders as institutional facts:

Consider for example a primitive tribe that initially builds a wall around its territory. The wall is an instance of a function imposed in virtue of sheer physics: the wall, we will suppose, is big enough to keep intruders out and the members of the tribe in. But suppose the wall gradually evolves from being a physical barrier to being a symbolic barrier. Imagine that the wall gradually decays so that the only thing left is a line of stones. But imagine that the inhabitants and their neighbours continue to *recognize* the line of stones as marking the boundary of the territory in such a way that it affects their behaviour. [...] The line of stones performs the same *function* as a physical barrier but it does not do so in virtue of its physical construction, but because it has been collectively assigned a new *status*, the status of a boundary marker (Searle 1995, 39-40; emphasis in the original).

Smit, et al. (2011, 8-9) offer the following alternative account of borders as institutional objects. They ask us to imagine two individuals who I will name, for the sake of consistency, Ana and Bob. These two individuals are lost on a desert island. After a dispute, Ana warns Bob that if he enters this half of the island she will beat him up. To give some substance to her words, Ana simultaneously points to two rocks on opposite sides of the island, cutting the island in half. Bob angrily replies to Ana that the same will be true for her if she enters his half of the island. Assume that the threats are credible. Given this assumption and provided that incentives to ignore the credible threats are insufficient for both individuals, then the invisible line drawn by the two rocks will actually function as a border. According to the authors:

It would be difficult to deny that a border had been set up on the island. But this has happened in a way that violates the essence of Searle's view. Firstly, the requirement for collective intentionality has not been met, as *all the relevant thoughts and claims can be expressed using the singular 'I'*. Secondly, nowhere is reference needed to any irreducibly social facts, objects, or properties. Both

actors can understand the situation fully by using concepts like ‘line’, ‘crossing’, ‘probability of getting beaten up’, etc. [...] We need nothing beyond an understanding of the incentives and beliefs of the two actors in order to grasp the situation fully (Smit, et al. 2011, 9, emphasis in the original).

The authors’ account makes the creation of the border depend on the effectiveness of incentives. Whatever the source of the incentives, as soon as they actually deter Ana and Bob to cross the imaginary line, a border has been created. In this case, contrary to what Searle argues, collective intentionality is not constitutive of the creation of the border.

I think that the authors’ argument, though perfectly in line with a standard rational choice or game-theoretic account, is partially misguided. It fails both to properly state Searle’s approach and to acknowledge the hidden epistemic requirements for their explanation to be valid. I will not delve into the details of the first point. It is sufficient to note that the authors appear to underestimate the role of language in Searle’s theory of institutional facts. Searle has forcefully argued that there cannot be institutional facts and institutions without language, i.e., that language is constitutive of the institutional reality.<sup>14</sup> A language is a set of symbolic devices that represent something beyond themselves. A language helps to convey meaning because it consists of symbolic devices that are *publicly* understandable.

Contrary to what Smit, et al. seem to presuppose, the institutional object (or fact) ‘border’ is tied to the existence of a linguistic device because the individuals must be able to use symbols to be able to form an intentional state with the border concept as a part of the propositional content. Or, in other words, individuals need to be able to represent what a border is before being able to desire or not to cross the border. Moreover, to define the concept of border, Ana and Bob will need other concepts such as ‘line’, ‘rock’, ‘to cross something’, and so on. To use Searle’s terminology, a border is a ‘language-dependent fact’ because it rests on a ‘language-dependent thought’, a thought that it is impossible to have without a language. Obviously, Ana and Bob do not need a language in Smit, et al.’s story; the authors *assume* that the border appears as soon as Ana and Bob, for whatever reason, do not cross an invisible line. The imaginary conversation in their story

---

<sup>14</sup> See, in particular, Searle 1995, 59-78.

is purely rhetorical, Ana and Bob could be language-less animals and it would still be appropriate to argue that a border has emerged.<sup>15</sup>

This leads to my second point. What the authors propose as an alternative account for institutional facts is a standard game-theoretic explanation. What I have said in the preceding two sections must suggest that they are wrong to conclude that their account is exempt from assuming any kind of collective intentionality. The ‘property game’ is easily modified to fit the border story. As soon as Ana has uttered the sentence ‘If you enter this half of the island I will beat you up’ while drawing a fictive line and Bob has answered similarly, the players have fully designed a set of possible states of the world  $\Omega$  where each state  $\omega$  is defined by the relative positions of Ana and Bob on the island.

Ana and Bob sentences are examples of performative illocutionary acts which Searle calls *declarations*: the propositional content of the acts (the semantic content of the sentences) is brought into existence by the very performance of the act. By uttering their (credible) threats through a linguistic device, Ana and Bob are doing three things at the same time: they partition  $\Omega$  into several states  $\omega$ ; they design their respective possibility sets  $P_i$  and their associated knowledge partitions; and they define an institution-as-correlated-equilibrium  $N(E) = \phi(\omega)$ . As I have argued above, all of this is common knowledge as soon as Ana and Bob have a common understanding of the situation, thus allowing us to assume mutual accessibility and symmetric reasoning. Obviously, this common understanding is permitted because Ana and Bob form a community constituted by the minimal fact that they share an institution, i.e., a language. If my argument that common understanding is analyzable as a form of collective intentionality is correct, then Smit and his co-authors are wrong when they claim that their account is free from any form of collective intentionality.

A possible answer to this argument is that, since language is unnecessary in the authors account, we can still assume that Ana and Bob progressively *learn* to not cross the line, with a border emerging as a result. In this type of evolutionary explanation, we do not need the epistemic apparatus beyond really basic assumptions

---

<sup>15</sup> Note that I am not arguing that Smit, et al. are necessarily wrong to define an institution as something that may emerge or exist without a language. Nonetheless, first, note that this is not how Searle defines an institution. Second, if we assume that collective intentionality depends on the existence of a language, then to define institutions as language-independent objects is clearly problematic. This last point is, however, an empirical issue.



regarding the learning mechanism (such as reinforcement learning, for example) because agents are assumed to be cognitively unsophisticated. This type of evolutionary framework works well to account for the evolution of animal behaviour.

In my opinion, ‘spontaneous order’ types of explanation are perfectly reasonable, in particular to account for the emergence of institutions. But they complement, rather than compete with, *eductive* explanations such as the one I have developed here. Eductive explanations are more relevant to account for the artificial construction of some low scale institutions (such as a business firm) and also to understand how widely acknowledged institutions are maintained through times.

### CONCLUDING REMARKS

In this article, I have presented a framework for introducing collective intentionality into game theory. Though I sympathize with the recent theories of team reasoning arguing for the possibility of team preferences, I think it is rather at the level of the formation of beliefs and intentions that collective intentionality plays a significant role. Basically, what I have been arguing is that when players have a common understanding of the situation, then they have common priors. Institutions can then be defined as correlated equilibria on which players can coordinate.

I have suggested that the common understanding concept embodies two epistemic requirements that must be added to standard Bayesian decision theory (Gintis 2009): mutual accessibility and symmetric reasoning. When common understanding obtains, players are able to develop a form of collective intentionality that shares many features with John Searle’s account. As a result, economists and in particular game theorists should not consider economic theory of institutions as competing with Searle’s theory of institutional facts. Indeed, the argument I have made in this paper suggests that game theory needs the Searlian perspective to provide a full explanation of institutions.

### REFERENCES

- Aumann, Robert. 1976. Agreeing to disagree. *The Annals of Statistics*, 4 (6): 1236-1239.
- Aumann, Robert. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55 (1): 1-18.
- Aumann, Robert, and Adam Brandenburger. 1995. Epistemic conditions for Nash equilibrium. *Econometrica*, 63 (5): 1161-1180.

- Bacharach, Michael. 2006. *Beyond individual choice: teams and frames in game theory*. Princeton: Princeton University Press.
- Bardsley, Nicholas. 2006. On collective intentions: collective action in economics and philosophy. *Synthese*, 157 (2): 141-159.
- Binmore, Kenneth. 1998. *Just playing: game theory and the social contract II*. Cambridge (MA): MIT Press.
- Binmore, Kenneth. 2007. *Playing for real: a text on game theory*. Oxford: Oxford University Press.
- Binmore, Kenneth. 2009. *Rational decisions*. Princeton: Princeton University Press.
- Bowles, Samuel, and Herbert Gintis. 2011. *A cooperative species: human reciprocity and its evolution*. Princeton: Princeton University Press.
- Chwe, Michael S.-Y. 2003. *Rational ritual: culture, coordination, and common knowledge*. Princeton: Princeton University Press.
- Cubitt, Robin, and Robert Sugden. 2003. Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. *Economics and Philosophy*, 19 (2): 175-210.
- Gintis, Herbert. 2009. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton: Princeton University Press.
- Hakli, Raul, Kaarlo Miller, and Raimo Tuomela. 2010. Two kinds of we-reasoning. *Economics and Philosophy*, 26 (3): 291-320.
- Hédoin, Cyril. 2012. Linking institutions to economic performance: the necessity of macro-structures in micro-explanations. *Journal of Institutional Economics*, 8 (3): 327-349.
- Lewis, David. 2002 [1969]. *Convention: a philosophical study*. Oxford: Blackwell Publishing.
- Searle, John R. 1990. Collective intentions and actions. In *Intentions in communication*, eds. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. Cambridge (MA): MIT Press, 401-415.
- Searle, John R. 1995. *The construction of social reality*. London: Penguin Books.
- Searle, John R. 2010. *Making the social world: the structure of human civilization*. Oxford: Oxford University Press.
- Smit, J. P., Filip Buekens, and Stan du Plessis. 2011. What Is money? An alternative to Searle's institutional facts. *Economics and Philosophy*, 27 (1): 1-22.
- Sugden, Robert. 1991. Rational choice: a survey of contributions from economics and philosophy. *The Economic Journal*, 101 (407): 751-785.
- Sugden, Robert. 2000. Team preferences. *Economics and Philosophy*, 16 (2): 175-204.
- Sugden, Robert. 2003. The logic of team reasoning. *Philosophical Explorations*, 6 (3): 165-181.
- Tuomela, Raimo. 2007. *The philosophy of sociality*. Oxford: Oxford University Press.
- Tuomela, Raimo. 2011. Searle's new construction of social reality. *Analysis Reviews*, 71 (4): 707-719.
- Tuomela, Raimo, and Kaarlo Miller. 1988. We-intentions. *Philosophical Studies*, 53 (3): 367-389.
- Vanderschraaf, Peter. 1998. Knowledge, equilibrium, and convention. *Erkenntnis*, 49 (3): 337-369.

**Cyril Hédoïn** is associate professor of economics at the University of Reims Champagne-Ardenne (France). His areas of specialization are in the philosophy of economics and institutional economics. He is currently working on the development of a non-individualistic understanding of game theory. He has published papers in the *Journal of Economic Issues*, the *Journal of Institutional Economics*, *Philosophy of the Social Sciences*, *Cahiers d'économie politique* (in French) and *OEconomia* (in French).

Contact e-mail: <cyril.hedoin@univ-reims.fr>