A Unified Social Ontology

by Francesco Guala and Frank Hindriks

Abstract:

Current debates in social ontology are dominated by approaches that view institutions either as rules or as equilibria of strategic games. We argue that these two approaches can be unified within an encompassing theory based on the notion of correlated equilibrium. We show that in a correlated equilibrium each player follows a regulative rule of the form 'if X then do Y'. We then criticize Searle's claim that constitutive rules of the form 'X counts as Y in C' are fundamental building blocks for institutions, showing that such rules can be derived from regulative rules by introducing new institutional terms. Institutional terms are introduced for economy of thought, but are not necessary for the creation of social reality.

Keywords: social ontology, institutions, regulative rules, constitutive rules, equilibria

1. Introduction

Social ontology studies the fundamental constituents of social reality. Although its subject is as old as philosophy itself – it dates back at least to Plato's theory of the state in *The Republic* – social ontology is still a disunified field with groups of researchers pursuing different projects and talking past each other at the expense of theoretical progress. In this paper we propose to remedy this state of affairs by presenting a general theory that unifies the currently dominant approaches to social ontology.

The social world is populated by entities such as norms, conventions, customs, laws, organizations, groups, identities and roles. Like other theorists we will refer to these seemingly diverse things using the generic term 'institution' together with its variants, like 'institutional role' and 'institutional fact'. The Constitution of the United States, for example, is an institution, while the President of the United States is an institutional role. And it is an institutional fact that Barack Obama is the forty-fifth President of the United States. An important task of social ontology is to explain the structure and functioning of institutions on

the basis of a relatively small set of principles or theoretical models.

Existing theories of social institutions can be classified in two broad categories depending on whether they conceive of institutions as *rules* or as *equilibria*. Theories within the equilibria approach view institutions as behavioural patterns or *regularities*. For example, Andy Schotter – a prominent game theorist and experimental economist – defines institutions as 'regularities in behaviour which are agreed to by all members of a society' (1981: 9). Such regularities 'can be best described as noncooperative equilibria' of strategic games (1981: 24), because out-of-equilibrium actions are unstable and are unlikely to be repeated in the course of many interactions.

The equilibria approach spans across the divide between philosophy and social science. The seminal theory in this tradition was proposed by David Lewis in a justly celebrated book on *Convention* (1969), and over the last three decades several other philosophers and social scientists have proposed equilibrium-based accounts of the emergence and persistence of social institutions.[1]

In spite of its explanatory achievements and its mathematical elegance, the equilibria approach has not been universally endorsed. According to an equally popular alternative, institutions should rather be conceived as *rules or constraints* that guide the actions of individuals engaged in social interactions. Douglass North – winner of the 1993 Nobel Prize for economics – for example, claims that

> Institutions are the rules of the game of society or, more formally, the humanly devised constraints that shape human interactions. […] They are a guide to human interaction, so that when we wish to greet friends on the street, drive an automobile, buy oranges, borrow money, form a business, bury our dead, or whatever, we know (or can learn easily) how to perform these tasks. (1990: 3-4)

The rules approach also cuts across the science/philosophy divide. The best-known proponent of a rule-based account of institutions in philosophy is John Searle, the author of a widely discussed book on *The Construction of Social Reality* (1995; see also 1969; 2005; 2010). In an article entitled 'What Is an Institution?' Searle claims: 'an institution is any system of constitutive rules of the form *X counts as Y in C'* (2005: 10; see also 1969: 51). Since

---

[1]   See e.g. Ullmann-Margalit (1977), Sugden (1986), Skyrms (1996, 2004), Calvert (1998), Young (1998), Aoki (2001), Vanderschraaf (2001), Binmore (2005), Bicchieri (2006).

Searle's constitutive rule theory has achieved great prominence in current philosophical debates, we will discuss it in depth in the course of this paper.[2]

The existence of two seemingly very different approaches raises some obvious questions. Are rule-based and equilibrium-based theories mutually incompatible, or do they simply focus on different aspects of social reality? Is one of the two theories more fundamental than the other, or do they provide ultimately incommensurable accounts? These questions will occupy us for the best part of this paper. Our contribution is ecumenical: we propose a unified account that is able to preserve the explanatory virtues of the two approaches, and we illustrate in detail how they relate to one another. We argue that institutions appear in the guise of rules or equilibria depending on the perspective that one takes ('internal' or 'external' to the game). We also demonstrate that constitutive rules like those advocated by Searle can be derived from simpler regulative rules expressed in conditional form (if X then do Y). And we show that regulative rules constitute correlated equilibria of coordination games.

The paper is organized as follows: section 2 briefly outlines the theory of conventions as correlated equilibria and shows that the distinction between the equilibria and rules approaches is spurious: institutions are perceived as rules when viewed from the perspective of the players, while they appear as equilibria when seen from an external, third-person perspective. Section 3 introduces Searle's constitutive rule theory and shows that it can be integrated within a general theory based on correlated equilibria and regulative rules. Section 4 extends the analysis to account for the normativity of institutions. Section 5 concludes with a summary and suggestions about the implications of the unified theory.

2. Conventions as Correlated Equilibria

In section 2.1 we present the equilibria approach to the analysis of institutions. Section 2.2 introduces the notion of correlated equilibrium, and in section 2.3 we argue that rule-based theories can be reformulated in equilibrium terms. This prepares the discussion, in the subsequent section, of the rules approach and its contribution to the analysis of institutions.

      2.1. Lewis' Theory of Conventions

---

[2]   On institutions as rules see also Parsons (1935), Knight (1992), Mantzavinos (2001), Hodgson (2006). Some hybrid theories cannot be classified neatly as either equilibrium- or rule-based; see for example Aoki (2007; 2011), Greif (2006) and Greif and Kingston (2011).

An equilibrium in game theory is a profile of strategies (or actions), one for each player participating in a strategic interaction. Each action may be described using a simple sentence of the form 'choose X' or 'do Y'. The defining characteristic – what distinguishes an equilibrium from other profiles – is that each strategy must be a *best response* to the action of the other players or, in other words, that no player can do better by changing her strategy unilaterally. If the others do their part in the equilibrium, no player has an incentive to deviate. This property is necessary but not sufficient for an equilibrium to become a behavioural regularity, because several profiles may be equilibria of the same game. This problem – known as the problem of 'equilibrium selection' – is at the origins of the seminal game-theoretic contributions to the analysis of institutions.

The first major breakthrough in the equilibria approach is due to David Lewis. Lewis (1969) proposed to model conventions as solutions to coordination games with multiple equilibria. His analysis focused on games with symmetric equilibria in which the players do not strongly prefer to converge on one rather than another solution. A classic example is the 'driving game': drivers do not particularly care about keeping right or left, provided everybody does the same. The theory however can easily be generalised to other cases, where the payoffs are asymmetric and the players have different preferences about the outcomes. Here we choose an example that has been discussed in some depth in the literature, and that provides a good story (or 'fable' if you prefer) about the origins of private property. The protagonists of our fictional story are two peoples living in the African savannah. As a homage to the great anthropologist Edward Evans-Pritchard, we have called them the Nuer and the Dinka:

> Many years ago, the Nuer and the Dinka settled in the valley of Sobat. The Nuer came from the north and the Dinka from the south, looking for green pastures for their cattle. Each tribe occupied as much grazing land as possible, until they arrived at the banks of the river Sobat. Moving their cows across the river would have been difficult, so each tribe grazed on one side of the river only (Figure 1a).

> Over the years the river progressively lost its water, due to changes in the region's climate, until at one point it became completely dry. Only a sandy line separated the areas occupied by the Nuer and the Dinka. The members of the two tribes could now easily trespass the old river's bed, and graze their cattle wherever they wished. But each piece of land now could be contested, and conflicts could easily escalate in outright war (Figure 1b).
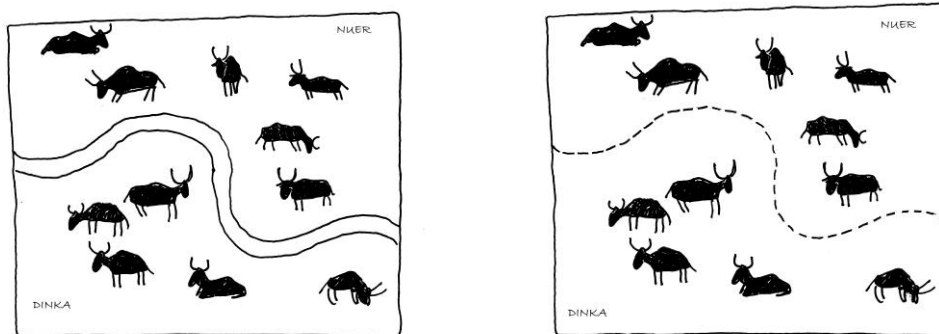
Figure 1 (a and b). Equilibrium in the valley of Sobat

The grazing game of the Sobat Valley can be represented in strategic form using a matrix known as 'hawk-dove' in biology, and 'chicken' in economics.[3] When the river dries up, the whole valley is up for grabs: for each piece of land now the Nuer and the Dinka ought to make a decision. In Table 1 the strategy G stands for 'graze', NG for 'not graze'. If they both decide to graze the same area, the members of the two tribes will end up fighting, which is the worst outcome for all (0, 0).[4] If they both abstain, they will not clash but will miss the opportunity to feed their cattle (1, 1). The best solution is to converge on one of the two equilibria in the top-right and bottom-left corners, where the member of one tribe grazes and the other lets him graze. But who is going to give way?

	G	NG
G	0, 0	2, 1
NG	1, 2	1, 1

Table 1. The grazing game (hawk-dove)

---

[3]   The use of this game to represent animal and human conflicts over contested resources goes back to Maynard Smith's (1982) work on evolutionary game theory. See also Sugden (1986) and more recently Gintis (2007).

[4]   In this paper we use the standard notation of game theory, unless otherwise indicated: the strategies of players are represented as rows and columns, the payoffs as numbers (the first one for the row player, the second one for the column player).

The grazing game is a problem of coordination with asymmetric equilibria, depending on who is going to give way. But since the players are perfectly identical, why should one of them accept a lower payoff? Notice that the only symmetric solutions here are not only inefficient, but are not even equilibria of the game. As a consequence we should expect some player to deviate unilaterally sooner or later.[5]

Our fictional story was designed so as to make a solution stand out as quite obvious. The story continues like this:

> Although it would have been very easy to trespass into the other tribe's territory, the Nuer and the Dinka preferred to avoid conflicts. The Nuer kept grazing on the north side, the Dinka on the south of the old river's bed (Figure 1b). The strip of sand could not *physically* stop raiders, but each tribe was happy to treat it as a border dividing their territories.[6]

Borders and territories are institutional entities. The Nuer and the Dinka have developed a rudimentary institution of property. But how can this institution be represented in game-theoretic form? Notice that borders or territories are not even represented in the grazing game (Table 1): they are extra-theoretic features that help find a solution to the coordination problem. As opposed to the actions in Table 1, the solution devised by the Dinka and the Nuer involves a new set of *conditional* strategies. Each tribesman, in particular, conditions his move (G or NG) on the position of the patch of land relative to the old river's bed. If it is north, the Nuer graze it, if it is south only the Dinka do it. This is similar to the solution of other problems analysed by Lewis (1969), such as the driving game that we have mentioned earlier. In that case, the drivers condition their choices on the *history* of play: if everybody has been driving on the right up until now, everybody will continue to drive on the right; if

---

[5]   Technically speaking, we are assuming a series of one-shot games with rematching (different players) at every round. The game is completely different – with more equilibria – if it is indefinitely repeated and the players have the possibility of building a reputation. Notice that in the one-shot setting there is also a mixed strategy equilibrium where each player chooses G or NG with probability 1/2. This equilibrium delivers expected payoffs of one unit each and is therefore inefficient. We ignore the mixed-strategy equilibrium for the time being, but will return to it briefly in the next subsection.

[6]   Searle (1995: 39-40) tells a similar fable about the emergence of borders and territories (1995: 39). He discusses an example in which a physical barrier – a wall – decays and evolves into a symbolic barrier. Our account is meant to explain how something like this might happen.

everybody has been driving on the left, it is in everybody's interest to keep doing the same. The only difference is that the conditional strategy in the driving game does not lead to a substantially different outcome than any of the two unconditional strategies ('keep right', 'keep left'). In the grazing game in contrast it creates a new behavioural pattern, for none of the unconditional strategies can deliver symmetric payoffs. This capacity – the capacity to create new outcomes – is an important feature of many institutions, as we shall see shortly.

2.2. Correlated Equilibria and the Emergence of Conventions

What kind of equilibrium can a convention be? Notice that there are two Nash equilibria and one convention in the grazing game. As a consequence, conventions cannot be Nash equilibria of coordination games.[7] Peter Vanderschraaf (1995) has shown that Lewis' conventions are *correlated equilibria*, a solution concept first studied by Robert Aumann in the 1970s. Correlated equilibria play an important role in the unified theory that we shall propose in this paper, so it is important to have an intuitive understanding of their characteristics. Since the formal models are somewhat complicated, we will provide here an informal account and leave interested readers to pursue the details in the technical literature.[8]

To grasp the idea of correlated equilibrium, it is useful to start from a hypothetical pre-conventional scenario. Suppose the Dinka and the Nuer are about to play the grazing game, and that (by hypothesis) no solution is salient. In such circumstances their only option is to choose randomly. The Nuer decide to toss a coin: if head comes up, they will choose G, if tails they will choose NG. The Dinka decide to do the same, and toss their own coin. Combining the probabilities that they obtain different results, there is fifty per cent probability that they will converge on one of the efficient solutions.[9] Their expected payoffs, unfortunately, are not larger than those that result from never grazing the contested land (1,

---

[7] This marks an important difference between our theory and the equilibrium accounts of Calvert (1995) and Greif and Kingston (2011).

[8] See Aumann (1974; 1987), as well as Vanderschraaf (1995; 1998; 2001) and Gintis (2007; 2009). Other equilibrium concepts may be used to articulate essentially the same idea, as in the literature on global games (Carlsson and van Damme 1993) and conditional games (Stirling 2011) for example. Signalling games (discussed by Skyrms 2010) are solved using refinements of correlated equilibria. In this paper we stick to the correlated equilibrium framework because it is the best known and most widely used.

[9] This is the mixed-strategy equilibrium mentioned in footnote 5.

1).

The coins in this example have been tossed *separately* and *privately*. Would it make any difference if, instead, the coin toss was a *single* and *public* event? Let us introduce a new character: following Gintis (2009) we shall call it the 'choreographer'. The choreographer tosses the coin and announces publicly: 'the Nuer graze' if head, or 'the Dinka graze' if tails comes up. The two players see the coin-tossing ceremony and, crucially, know that the other can see it too. Moreover, they know that the other knows (and so forth) that they both see the same ceremony: the outcome of the toss is common knowledge.

In such circumstances, it would seem reasonable to follow the advice of the choreographer. Each player, in other words, could condition her behaviour on the result of the toss, following the obvious strategy: 'choose G if the choreographer says so, otherwise choose NG'. If both players are confident that the other one is going to follow this strategy, then they are better off tossing a single coin publicly. The signals would be perfectly correlated, so the players could always coordinate on an efficient outcome, exploiting the results of the ceremony.

Solutions of this kind are correlated equilibria. A correlated equilibrium of a game $\mathcal{G}$ is a Nash equilibrium of a larger game $\mathcal{G}^*$ obtained augmenting $\mathcal{G}$ with the addition of new strategies. The new strategies prescribe actions conditional on the occurrence of an external event that is not part of the original game. They take the form 'if X then do Y', where X is a property of the correlation device. As Vanderschraaf has shown, Lewis' conventions are correlated equilibria that exploit earlier moves in the coordination game. The coin, in other words, is replaced by the history of play.

What would an augmented version of the grazing game look like? The introduction of an external device breaks the stalemate, creating a better solution than those that were previously available. The external device that the Nuer and the Dinka use to condition their strategies is the location of the contested piece of land with respect to the river Sobat. If the land is north of the river, the Nuer graze it; if it is south, the Dinka graze it. This is analogous to tossing a coin or using past behaviour as a guide for future action in the driving game. We construct an augmented grazing game adding two conditional strategies to the game in Table 1 (with 'N' for 'north', and 'S' for 'south'):

(i) G if N, NG if S;

(ii) G if S, NG if N.

Again let us assume for simplicity that in each instance there is a fifty per cent chance that the land lies north (or south) of the river. Notice that in the augmented game (Table 2) the outcome of the conditional strategies is the only symmetric Nash equilibrium in pure strategies (1.5, 1.5). The other two equilibria (G, NG) and (NG, G) are still there, and each player would in principle prefer to converge on the one that is most favourable to her. But conditioning on an external event provides a third stable solution that should be acceptable to both.

	G	NG	G if S, NG if N
G	0, 0	2, 1	1, .5
NG	1, 2	1, 1	1, 1.5
G if N, NG if S	.5, 1	1.5, 1	1.5, 1.5

Table 2. Augmented grazing game

2.3. Equilibria and Rules

Of course one may ask: why *this* kind of correlation? Why do the tribesmen pay attention to the river, rather than to some other element of the environment? Surely one can construct a thousand augmented games, corresponding to thousands of correlated equilibria that exploit different external events. 'Graze if sunny, do not graze if cloudy'; 'graze if Monday, do not graze if Tuesday': these are all potential correlated strategies. Why 'G if S, NG if N'? What is so special about it?

Schelling (1960) and Lewis (1969) have argued that a strategy profile may be selected in virtue of its salience. That the Nuer graze north and the Dinka graze south of the old river is salient in the Sobat Valley, because it has been the traditional arrangement for many years, before the river became dry. History creates a focal point for coordination. If no event or correlation stands out 'naturally', however, a group of ingenious players can try to make one of them *more* salient – by means of teaching, acculturation, and training. But there is no game-theoretic account of which correlations we naturally hook onto. Social ontology merges with social history, psychology, and biology at this point, and game theory must give way to

empirically informed models.[10]

Clearly the theory of correlated equilibria does not solve the mystery of salience. It does, however, solve an important puzzle of social ontology. Recall the question we started from: Are institutions *rules*, or *equilibria* of a game? We can now see that the answer is 'both': an institution may be considered as an equilibrium or as a rule of the game, depending on the perspective that one takes. The correlated equilibrium in the game of the Sobat Valley is the pair of strategies:

(i)   G if N, NG if S.

(ii)   G if S, NG if N.

From the point of view of an external observer, the convention that regulates grazing in the Sobat valley takes the form of a *regularity* that corresponds to the correlated equilibrium in $\mathcal{G}$ (or, which is equivalent, to the corresponding Nash equilibrium in $\mathcal{G}^*$). But each strategy in this profile also takes the form of a *rule* that dictates each player what to do in the given circumstances. The Nuer therefore will perceive the institution as a prescription to graze their cattle if the land is north, not to graze if it is south. And, mutatis mutandis, the same is true for the Dinka (G if S, NG if N). Notice that conditional strategies so far are normative only in the weak sense of instrumental rationality: given what the other players do, it is best to do one's part in the equilibrium (we will return to normativity again in section 4). Since the two strategies are formulated as rules, clearly the equilibrium is a set of rules – one for each player – that 'establish a stable structure to human interaction' (North 1990: 6).[11]

We should like to emphasise again that it would be impossible to conciliate these two views if we focused on the Nash equilibria of the original matrix (Table 1): the conditional strategies (rules) are not even part of this game. And they cannot be: there is no north/south correlating device in the original game. So it is correct to see the correlated strategies as *external* rules that help attain coordination in the original game (North 1990). But the convention is not a Nash equilibrium of the original game – it is a correlated equilibrium of

---

[10]   For some attempts to enrich equilibria accounts with empirical evidence about human cognition, see e.g. Sugden (1986), Mantzavinos (2001), Gintis (2009).

[11]   This incidentally explains why the proponents of rule-based accounts sometimes use game theory models to explain the functioning of institutions. The view that institutions are endogenous products of games, even though they may appear as exogenous to the individual players, is explicitly endorsed by Binmore (2005) and Aoki (2007).

the old game, and a Nash equilibrium of the *augmented* game. The contrast between rules and equilibria approaches is perhaps due to a failure to appreciate this distinction between equilibrium concepts. With the introduction of correlated equilibria both approaches are vindicated. Thus, we have achieved a unified view of social ontology.

But it would be a shame to stop here. The rules that we have been concerned with are *regulative* rules, rules that govern behaviour in strategic interactions. It remains to be seen whether our approach is consistent with theories that explicate institutions in terms of constitutive rules, that is, rules that enable the performance of certain actions. The intuitive idea is that it is not possible to perform actions such as playing chess or registering a limited liability corporation independently of the rules that constitute the relevant institutions. But we will see that these theories, which have attracted a lot of attention in the philosophical literature, do not provide a genuine alternative to the equilibrium-based account of institutions either. On the contrary, constitutive rule theories can also be incorporated in the unified theory outlined above, because they can be analysed in terms of regulative rules and equilibrium outcomes.

3. The Constitutive Rule Theory of Institutions

The most prominent constitutive rule theory of institutions has been proposed by John Searle (1995; 2005; 2010).[12] In this section we argue that Searle's account is compatible with the unified theory introduced in section 2. This is important, because Searle believes that his theory is radically different from rival theories such as the ones discussed so far, and makes use of terms and concepts that have nothing in common with those commonly used in the game-theoretic tradition. In particular the notion of a constitutive rule seems to be alien to the approach of economists and other social scientists who view institutions as behavioural regularities or equilibria. But we shall see that this impression is illusory: Searle' account can be reformulated in such a way as to make it compatible with equilibrium-based theories: constitutive rules play an ancillary role in the analysis of institutions, and the latter can be analysed entirely in terms of regulative rules.

---

[12]   See Tuomela (2007), Hindriks (2009, 2012, 2013) and Marmor (2009) for other versions of the constitutive rule theory.

3.1. Searle's Theory

The rules that appear in the unified theory prescribe specific actions in specific games. They are *regulative* rules, in the sense that they regulate the behaviour of individuals engaged in certain social interactions. In a number of books and articles, Searle has argued that regulative rules do not suffice to explain the complexity of social reality. Rules of a different kind, *constitutive rules*, are necessary for the creation and maintenance of social institutions. Searle has proposed a memorable formula to capture the underlying logic of all institutional facts. A constitutive rule stipulates that

X counts as Y in C,

where X is (or can be) a pre-institutional entity, Y is a status function and C refers to the domain of application of the rule. For example, putting the ball in the net (X) counts as scoring a goal (Y) in a game of European football (C).

Of course a single rule does not exhaust a complex institution like football. An institution is typically constituted by a large number of interlocking rules that jointly define and regulate the activities of individuals. As Searle rightly points out, the rules form constraints on the moves that the players can perform on the pitch, but at the same time they also create opportunities that otherwise would not exist. You cannot score a hat-trick or save a penalty if the institution of football does not exist. Most of the things that take place on the football pitch are *institutional facts*, and they would be impossible if the appropriate rules were not in place. In Searle's own words, 'institutional facts only exist within systems of constitutive rules' (1995: 28). As mentioned in the introduction, Searle goes as far as identifying institutions with (systems of) constitutive rules.

Searle's emphasis on the creative power of constitutive rules is backed up by his distinction between constitutive and regulative rules.[13] The criterion is intuitive: if a rule regulates an antecedent activity, it is a regulative rule; if it makes a new activity possible, then it is a constitutive rule. One of Searle's favourite examples is the game of chess: 'the rules are *constitutive* of chess, in the sense that playing chess is constituted in part by acting in accord with the rules' (1995: 28; see also 1969: 33-42).

---

[13] The distinction goes back to Anscombe's (1958) analysis of brute facts and Rawls' (1955) two conceptions of rules. Searle (1969) argued that, against a background of brute facts, constitutive rules serve to create institutional facts. See Ransdell (1971), Conte (1988), Lagerspetz (1995), Tuomela (1995; 2002; 2007), Hindriks (2009; 2012; 2013), and Morin (2011) for more recent discussions of the notion of a constitutive rule.

3.2. Constitutive Rules Deconstructed

The main goal of this section is to debunk the distinction between regulative and constitutive rules (see also Hindriks 2009, 2012). This may seem like a difficult task, for the grammatical forms of these rules appear to be very different. The difficulty is largely illusory however, as we shall see shortly. Let us start from Searle's basic formula for constitutive rules:

X counts as Y in C.

The phrase 'counts as' in Searle's locution indicates that the people in the context at issue accept X as Y, or agree that X is Y.[14] The rule then has the following structure:

X is Y in C.

X, Y, and C can be regarded as predicates that feature in a universal generalization: any entity that is X and that is in context C has status Y (Hindriks 2009: 263). Searle has recently suggested that it would be more accurate to replace the X-term with the locution 'for any x that satisfies conditions p' (2010: 99). This gives us a formula of the following kind:

if P, then X is Y in C.

For example: if the striker is not in off-side position, then putting the ball in the net is scoring a goal in a game of football.

What remains to be shown is that Y involves actions, in such a way as to make the distinction with regulative rules all but trivial. In this section, we begin to argue that statuses feature actions implicitly. Furthermore, we make this explicit introducing 'status rules' that specify the practical significance of statuses. On our view, Searle's basic formula is best seen as an elliptic statement of the structure of constitutive rules. This is important, because once status rules are seen as proper parts of constitutive rules, it becomes apparent that any regulative rule can be transformed into a constitutive rule, and vice versa, by adding or eliminating

---

[14] Most proponents of the equilibria approach invoke no attitudes other than individual preferences and expectations (see, however, Bacharach 2006; Bardsley 2007; Gold and Sugden 2007). Many of those who endorse the rules approach believe that collective acceptance involving commitment is required instead (Gilbert 1989; Searle 1995; Tuomela 1995; 2002; 2007; note that Searle 2010 has recently switched sides). Since this issue is orthogonal to our concerns, however, in this paper we remain uncommitted as to what kind of attitudes are involved in institutions.

theoretical terms. This transformation is the basis of our challenge to the distinction between the two kinds of rules.

Consider once again the scenario that we introduced in section 2: in the original story the Nuer and the Dinka had found a convenient solution to the problem of grazing in the Sobat Valley. Using the river's bed as a correlation device, the Nuer grazed by convention the land that lies north of the river, and the Dinka the land that lies south. Now whenever a Dinka and a Nuer man look at a piece of land, they immediately notice its location and apply their conditional strategies – they play their part in the following correlated equilibrium:

[R]   (i) Graze if the land lies north, do not graze if it lies south (G if N, NG if S).

        (ii) Graze if the land lies south, do not graze if it lies north (G if S, NG if N).

These statements are *regulative rules* in Searle's terminology. They describe pre-existing activities, regularities of behaviour that have emerged in the past for whatever reason. Note that the existence of conventions does not require an explicit formulation of the rules that, as a matter of fact, define the correlated equilibrium of the game (the equilibrium could for instance be sustained simply by imitation). But the main point is that conventions of property are institutions, and if such institutions only require the existence of regulative rules, our analysis is inconsistent with Searle's theory.

Of course one may restrict the notion of institution so as to satisfy Searle's criterion. Conventions would be excluded by definition because they only involve regulative rules, and one might think that it is a fair price to pay for a philosophically sound and sharp taxonomy of social entities. But such a move is bound to fail. It is in fact possible to derive constitutive rules from regulative rules via the introduction of institutional terms; conversely, constitutive rules can be turned into regulative rules via the elimination of these terms. Since institutional terms are eliminable at no cost, *constitutive rules are, at roots, just regulative rules dressed up in institutional language*. Searle's distinction breaks down, which means that his theory can be encompassed within a unified social ontology based on correlated equilibria and regulative rules.

Notice that the terms that appear in [R] refer to specific actions ('graze on the north side', 'do not graze on the south side', and so forth) that apply to this particular game. Nothing however prevents one from inventing a more general concept that *implies* the actions in such circumstances. Let us introduce the term 'property', to denote all the land where the members

of a tribe can graze their cattle. Since we shall use a somewhat artificial concept, much simpler than our full-blown notion of property, we add the star symbol (*) to mark its peculiarity. A (partial) definition of the term property* is provided by the following *status rule*:[15]

[S] If a piece of land is Nuer's property*, then only the Nuer graze it.

The status rule introduces the new term property*, and (partially) defines it in terms of previously understood terms like land, grazing, and the Nuer people. The *application* of a new term requires criteria however. When is it appropriate to say that something is Nuer's property*? This is determined by a *base rule* like this:

[B] If a piece of land lies north of the river, then it is Nuer's property*.

And here comes the crucial step: the conjunction of a status and a base rule [S+B] forms a *constitutive* rule [CR] in Searle's sense.[16] For example:

[CR] If a piece of land lies north of the river, then it is Nuer's property*, and if a piece of land is Nuer's property* then only the Nuer graze it.

Constitutive rules, thus, have the following grammatical form:

If P, then X is Y, and if Y then Z.

At this point one may complain that the [CR] formula merely plugs a regulative rule into Searle's original formula. The base rule [B] has exactly the same grammatical form of Searle's XY formula (if P then X is Y). So why should we invent a new name for it, and use the "constitutive" label for the augmented XYZ formula instead? The reason is that Searle's version of the constitutive rule formula is elliptic. Unless we say what Y stands for, it does not mean much. The formula must be made explicit by specifying the content of Y, and once that has been done, we can see that the full constitutive rule has the XYZ grammatical form. This is a *substantial,* not just a grammatical point, and for this reason we find it helpful to use

---

[15]   The term 'status rule' was introduced by Hindriks (2009, 2012) and is shaped after Searle's *status function*: a non-institutional entity (a piece of land in this case) acquires a new status (it is Nuer's property) in virtue of a rule that attributes a new function to it (since statuses and powers are related in Searle's view, we explicate statuses in terms of normative powers rather than functions; see section 4).

[16]   In this respect our terminology differs from that used by Hindriks (2009; 2012; 2013). We use the term 'base rule' for what he calls a 'constitutive rule' reserving the latter term for the conjunction of a status and a base rule.

the [CR] label for the full rule. The claim is that without a Z-term the constitutive rule does not mean anything at all.

Decomposing constitutive rules into base and status rules helps appreciate that Searle's distinction between regulative and constitutive rules does not capture a substantial ontological divide. It is clear that introducing the term property* does not add anything new to the pre-existing rules in [R], given the definition [S]. Another way to put it is to say that property* existed before the constitutive rule was formulated explicitly, so the rule only introduced a new term to refer to the practices of the people who live in the Sobat Valley. Whether a regulative rule is in fact transformed into a constitutive rule or not is tangential to the existence of institutions. We refer to this view as the 'transformation view of constitutive rules'. [17]

At the core of the transformation view lie the following two claims: institutions may be described entirely in terms of rules that regulate pre-existing activities; and regulative rules can be extended theoretically to create constitutive rules and generate new institutional terms at no cost. The question arises what the value might be of introducing constitutive rules. Our answer is twofold. The practical advantage of constitutive rules is that they are conducive to economy of thought. Their theoretical advantage is that they make the ontology underlying regulative rules explicit. In the next sub-section we elaborate on this answer by explaining more precisely what the theoretical extension amounts to.

3.3. Institutions as Theoretical Terms

A system of constitutive rules is a theory of sorts. Theoretical terms like property* refer to

---

[17]   Nuel Belnap's (1993) criteria for rigorous definitions may help us see the point more clearly. Belnap argues that, in order for a definition to be rigorous, it should satisfy the criteria of eliminability and conservativeness. The criterion of eliminability requires 'that the defined term be eliminable in favor of previously understood terms', and the criterion of conservativeness demands 'that the definition not only not lead to inconsistency, but not lead to anything – not involving the defined term – that was not obtainable before' (Belnap 1993: 117). In other words, a definition of a term is rigorous if we can do without it, and if, in case we choose to use it nevertheless, it does not introduce anything that is qualitatively different from what can be expressed by only using terms previously understood. We will say that [CR] is a *conservative transformation* of [R], because the definition of the institutional term property* is rigorous in this sense. Given the criteria of eliminability and conservativeness, conservative transformations do not involve qualitative changes of the theory at issue.

entities that exist quite independently of the theory itself – in this case they refer to a list of strategies or regulative rules. One reason why it is useful to have property* in one's vocabulary is that it spares the trouble of listing all the actions that comprise the equilibrium between the Nuer and the Dinka. The theoretical term property* however can in principle be dispensed with, by formulating the theory entirely in regulative language.[18] In fact this happens with lots of institutions to which we never bothered assigning a name. We do not have a special term for the institution of driving on the left-hand side of the road, for example. We could call this institution 'leftism' if we wanted to, but as a matter of fact we do not feel the need to introduce a new term in our language for this convention.

Why is that? A plausible explanation is that this institution is too simple to deserve a name. The institution of driving on the left is pretty much exhausted by a single regulative rule (if you are, for instance, in Japan or Britain, then drive on the left). Theoretical terms are useful if they achieve economy of language and thought, and in fact we usually assign names to institutions, like property or marriage, that are more complicated than driving on the left. Consider property again: although in our fictional example a single action (to graze) is mentioned in the status rule [S], in a realistic context the list should be much longer: there are many things that the Nuer and the Dinka could do within their territories without unsettling the coordination equilibrium. They could dig, plough, harvest, even sell their land. Property rights typically include the right of exclusive use, residual claim, and the right to transfer.[19]

The upshot of all this is that Searle's distinction between regulative and constitutive rules is not essential for capturing the nature of institutions. Constitutive rules are regulative rules refurbished by means of a new term. Since the term names an institution, institutions are (systems of) regulative rules. Because the regulative rules may have existed for years prior to the introduction of the new term, stating the constitutive rule does not necessarily create anything new. Such a linguistic innovation is not needed for the institution to exist.[20]

Of course this does *not* mean that language is useless. On the contrary, the amazing

---

[18]   There is a close analogy between the function of institutional terms and the role of theoretical terms in science (see e.g. Lewis 1983a; 1983b), which underlies the claim that constitutive rules are theories of sorts (Hindriks 2009, 2012).

[19]   The conception of property as a 'bundle of rights' is very common in economics; see e.g. Demsetz (1967) and, for critical discussion, Penner (1996).

[20]   This result may be used to argue that language is not a fundamental building block of institutional reality. A proper discussion of this claim would require a separate paper, but see for example Hindriks (2011).

complexity of human societies can arguably be achieved only by communities of individuals who have developed sophisticated means of communication and representation. But we can account for this pragmatic role of language without postulating ontological powers it does not possess. Linguistic utterances are just another, extremely flexible and effective device for coordination and equilibrium selection.

In the fictional case of the Nuer and the Dinka, we can easily imagine an alternative story, where equilibrium selection is performed by means of a linguistic utterance. Suppose that the Nuer and the Dinka found no river, when they first entered the Sobat Valley. Suppose that the elders of both tribes met in the middle of the plain and tried to agree on a division of land. After a lot of haggling they stood up, called a plenary meeting, and declared:

[R'] All the land that lies north of this point is grazed by the Nuer, the land south is grazed by the Dinka.

The rule in this case would have not described a pre-existing regularity, for the equilibrium solution had just been identified theoretically – 'predicted' if you wish – by the elders. Here we have a clear example of the causal power of language: if the elders exert some authority over the other members of their tribes, or they are simply able to create a focal point, then stating the new rule [R'] is likely to facilitate convergence on that specific equilibrium of the game. The elders work like an equilibrium selection device – like choreographers – by making a public statement. Uttering the rule makes one equilibrium salient.[21]

Thinking about a rule and agreeing to follow the rule can affect the range of things that people do. By conditioning their actions on the position of the river's bed, the Nuer and the Dinka bring about a new equilibrium that did not exist before. When Searle says that some rules 'create the very possibilities of certain activities' (1995: 27) he might have something like this in mind. This is not a novel insight though: social scientists have been aware for decades that linguistic statements – especially predictions – may work as self-fulfilling prophecies.[22] If we believe that the other members of the community have a propensity to choose the strategy that has been made salient by the statement, and we have a preference for conformity (because we are playing a coordination game), then we will also choose that

---

[21]   Since this requires sophisticated cognitive capacities, like the capacity to represent the content of the rule and store it in one's memory, it is not surprising that humans are the only animals who construct complex institutional architectures.

[22]   See e.g. Merton (1948), Simon (1957).

strategy. If the statement is a prediction, then our behaviour will verify the prediction. If the statement is a rule like [R'], then our behaviour will conform to the rule. Additionally, if the rule is a constitutive rule, we will have the impression that a new social institution (like property*) has come into existence by mere stipulation. A proper ontological analysis makes clear that it is only an impression: language is one among many coordination devices, and has no more creative power than a coin toss or any other event the players may use to coordinate their decisions.

4. From conventions to norms: adding normative power

We have used the term property* (with an asterisk) to signal that this is a toy version of the actual institution of property. As we said, the institution of private property is not linked to one kind of action only – if it were, the term would be quite redundant. But property* differs from property also in another respect: the statuses of real institutions are usually expressed in normative or deontic language. For example: a piece of land may be Nuer's property even though the Nuer do not actually graze it. The point of it being their property is that the Nuer *can* graze it if they want, or *may* do so. Whether they do it or not depends on various contingent circumstances. For example, the Nuer may decide that they prefer to let the grass grow, or plough the land, or rent it out to an oil company. They are permitted to do all these things, if it is their property, while the Dinka are not.

According to Searle, 'deontic powers' are an essential feature of institutions:

> The simplest test for whether a phenomenon or fact is genuinely institutional is to ask, Does its existence imply deontic powers, powers such as those of rights, duties, obligations, requirements, and authorizations? (Searle 2010: 91; see also 8-9)

But what are these powers, and can they be accounted for within the unified theory that we propose in this paper? In this section we follow an established tradition in the theory of choice, showing that normative power can be modelled as a cost that modifies the payoffs of a game. This allows the extension of the theory outside the domain of coordination games, to include also situations where the players have selfish incentives to deviate from a socially superior equilibrium, or where they are uncertain about the real structure of payoffs, or about the motives of the other players. As we will see shortly, this extension does not involve any commitment to a specific theory of normativity. Since deontic powers may have several sources and may be sustained by different mechanisms, the unified theory should better be

neutral about them.

First, however, let us see how these powers may be represented formally: what happens if a normative element is added to the rules discussed so far? Consider the following transformation of [R'] into two separate normative rules, one for each tribe:

[R''] If a piece of land lies North of the river, then the Nuer *can* graze it.

[R'''] If a piece of land lies North of the river, then the Dinka *cannot* graze it.

These statements or rules are expressed in normative terms (as used here, 'can' is equivalent to 'may' or 'is allowed'). [R''] specifies a permission, and [R'''] a prohibition. As a permission is simply the absence of a prohibition, a *norm* can be seen as a convention augmented with deontic force imposing extra constraints on the behaviour of some (but not necessarily all) players.

When philosophers claim that a rule has normative force they imply that people have reason to follow it (Black 1962). In this vein Searle argues that deontic rules provide a special type of reason to act:

> Deontic powers have a unique trait, [...] I think, uncommon and perhaps unknown in the animal kingdom: once recognized they provide us with reasons for acting that are independent of our inclinations and desires. (Searle 2010: 9)

Someone might, of course, happen to have a desire to abide by a rule. The point, however, is that she still has reason to do so in the absence of such a desire. Furthermore, she also has a reason to follow the rule if she has a conflicting desire. Norms usually give reasons *not* to pursue broadly self-interested goals, so conflicting desires are not uncommon.

A convenient way of modelling reasons against violating rules is in terms of costs. As long as the term is understood in a sufficiently broad way, we can say that norm compliance carries costs while conforming to (non-normative) conventions does not (Crawford and Ostrom 1995). For example, a littering norm imposes the cost of looking for garbage bins. A norm against cheating imposes the opportunity cost of not having extra-marital affairs, and so forth. Such opportunity costs call for compensation (counter-reasons) in order to be overridden. And counter-reasons may be represented as negative incentives – other costs, effectively – that deter norm violations. In general, the normative element of a rule refers to the system of incentives that supports an equilibrium when the players may have selfish reasons to deviate. The unified theory is compatible with various incentivising mechanisms: people may tend to

feel some desire to avoid resentment simply because they want to preserve the goodwill of others. But they may also realize that resentment may lead to sanctions. Crawford and Ostrom (1995) distinguish between internal and external mechanisms, corresponding roughly to a distinction between 'internalized' and 'externally sanctioned norms'. Both are represented by a delta parameter to capture the costs involved in violating a norm (1995: 587).

Representing normative power by means of delta parameters (costs) facilitates the extension of the unified theory beyond the realm of coordination games. Many social theorists have noticed that some institutions improve the performance of players in games where there is a positive incentive to deviate from the socially optimal rule.[23] A classic case is the famous prisoner's dilemma game (Figure 2(a)): defecting in the (one-shot) prisoner's dilemma is the only rational strategy, because D dominates C. Instead of multiple equilibria – as in the coordination games examined so far – there is only one equilibrium in the prisoner's dilemma (DD). Augmenting the game by means of conditional strategies does not help in this case, so there is no way to solve a prisoner's dilemma using an external correlation device. Defecting strictly dominates rule-following. By means of a suitably large delta, however, one can transform the prisoner's dilemma in a game with two equilibria, CC and DD. Normative rules thus can turn a dilemma of cooperation into a coordination problem: norms change the games that people play. And, of course, they can work as coordination devices (choreographers) in the coordination game that has just come about.

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

(a)

	C	D
C	2, 2	0, 3-δ
D	3-δ, 0	1, 1

(b)

	C	D
C	2, 2	0, 0
D	0, 0	1, 1

(c)

Figure 2. Transforming a prisoner's dilemma into a hi-lo game

Suppose there is a rule in the population that says 'if the other player cooperates, then you

---

[23]   See e.g. Ullmann-Margalit (1977), Sugden (1986), Crawford and Ostrom (1995), Binmore (2005), Bicchieri (2006), Gintis (2009).

ought to cooperate, otherwise defect', and the rule has normative force. This means that an extra cost (delta) must be subtracted from the payoffs, as in Figure 2(b). The payoffs in these two games may be interpreted in various ways: one possibility is to take the numbers in the original game (a), as representations of objective (material) payoffs. In the new game (b), then, the numbers may be subjective (utility) payoffs. Many game theorists frown at this distinction, though, and prefer to use the utility interpretation in all cases. In this case the modified payoffs of Figure 2(b) may result from new information that has become available to the players, for example when they discover that a certain normative rule is in place or that a system of punishments has been set up to deal with those who transgress it. In any case the deltas represent the force of normative rules, taking a given (pre-normative) game as a benchmark.

Depending on the force of the rule the second game may turn out to be quite different from the first one. If delta is at least as large as one unit of material payoff, then the prisoner's dilemma is transformed into another simple game, where DD and CC are both equilibria. If $\delta=3$, for example, we obtain the game of Figure 2(c) (called "hi-lo").

Norms then "solve" dilemmas of cooperation only in a peculiar sense. There is no way to escape the disturbing conclusion that players ought to defect in the standard prisoner's dilemma, without changing the rules of the game. The only 'solution' is to change the game itself, and this is precisely what normative powers can do. They can create new equilibria introducing costs that make defection unattractive, at least within a certain range of payoffs.

In so doing they also fulfil two key functions of institutions, highlighted by North (1990) and many other social scientists: institutions stabilize behaviour and make it more predictable in situations of uncertainty; and they create new equilibria that did not exist before, by changing the payoffs of a game. If a sufficiently strong norm is in place, we can afford to ignore small variations in our material incentives and in the incentives of the other players. Following a simple rule (if X then do Y) then may be profitable regardless of whether the strategic situation we are facing is a coordination game or a dilemma of cooperation.[24] Thus, norms explain the persistence of institutions, or how they can remain in force in the face of incentives to deviate. But they also play a role in the emergence of *new* equilibria, when they are introduced ex novo. A central authority – like a government or a recognized leader, can

---

[24] The study of 'complex games' with variable payoffs is in its infancy, but see e.g. Zollmann (2008), Wagner (2012).

re-design a game with 'bad' equilibria, such as a prisoner's dilemma. New rules may be introduced by decree – as the elders did in the alternative story of the Nuer and the Dinka that we told at the end of section 3. If the decree is supported by credible formal and informal sanctions, the players will recognize that the game has changed and that a new equilibrium has been created. Searle's insight that institutions make new types of behaviour possible is fully vindicated in the unified account.

Notice that the unified theory only provides a formal apparatus to represent deontic powers, but is not committed to any specific theory of normativity. Formulating a rule in deontic form only highlights the existence of some mechanism that induces the relevant players not to pursue their selfish ends. In ordinary language 'ought' statements are ambiguous regarding the nature of such mechanisms – whether they are internal or external to the individuals, for example [25] – and similarly the unified theory is neutral about what normativity is or where it comes from. Far from being a defect, we think that this is just as it should be. Normativity is one of the thorniest issues in contemporary philosophy, and it would be foolish to make a theory of institutions depend on a specific account. Some philosophers and social scientists believe that normativity can be analysed in terms of mutual beliefs and the feeling of resentment that we experience when our expectations are frustrated (Lewis 1969; Sugden 1998); others believe that normativity requires a stronger notion of collective agreement or joint intention (Gilbert 1989); some philosophers and social scientists argue that normativity depends on emotions (Frank 1987; Gibbard 1990; Nichols 2004); and still others believe that normativity has to do with the possibility of justifying our actions by means of rational arguments (Raz 1999; Skorupski 2010). Whether any of these accounts is able to explain normativity in a satisfactory way is an open issue that we do not need to settle here. And it is not even clear that we ought to choose among these theories: normativity is likely to have different sources and many facets, so more than one account is probably right.[26]

5. Conclusions

---

[25]   When we tell our kids that they ought to tidy up, for example, we may be appealing to their internalized moral principles, or threatening sanctions, or both.

[26]   Turner (2010) provides an excellent discussion of the social science approach to norms, and a critique of the futile search for "true" normativity. From a theoretical perspective, our approach is similar to Gintis' (2009); unlike Gintis, however, we are not committed to the claim that the human predisposition to follow norms is a product of natural selection.

The discussion of the normative dimension of institutions concludes our argument for the thesis that the two main traditions within social ontology – the rules and the equilibria approaches – can be unified into an overarching theoretical framework. The framework is entirely general. Recall that the notion of a correlated equilibrium played a central role in our argument that the equilibria and the rules approaches are complementary to one another. This argument does not depend on anything that is unique to our guiding example, the Nuer-Dinka grazing game, but extends to any institution that can be analyzed in terms of the notion of a correlated equilibrium. Institutions such as marriage and money can be easily represented as solutions of coordination problems. And even prisoner's dilemma situations, as we have seen, can be turned into coordination games if a system of normative powers is in place. The key consideration is simply that conditional strategies can be seen as rules, a point that gains plausibility once the normative dimension of institutions is taken into account.

A similar argument can be developed in relation to the transformation view of constitutive rules. We illustrated it for the case of Nuer-property*, which admittedly is a toy example. Actual property goes beyond the right to use and includes, for instance, the right to transfer. There is nothing in the method of transformation that we have proposed that is unique to the right to use land for grazing. Any regulative rule can be transformed into a constitutive rule by first introducing an institutional term and then using it to transform the antecedent of the regulative rule into a base rule, and its consequent into a status rule. In order to accommodate the right to transfer, the base rule of Nuer property* can remain the same: If a piece of land lies north of the river, then it is Nuer's property**. The status rule has to be generalized so as to include the new right: Nuer can transfer land that is their property* as well as graze it. As its steps are perfectly general, the transformation view also applies to institutions other than property. The base rule of marriage, for example, explicates the conditions people have to meet in the context at issue in order to have the status of a married person. Its status rules specify the normative powers that come with this status, and normative powers can be expressed as regulative rules.[27] The upshot is that, if an institution can be explicated in terms of a regulative rule, it can be redescribed in terms of a constitutive rule.

What is the added value of the theory that we have laid out in this paper? First, our unified theory makes it possible to appreciate the commonalities and differences between theories

---

[27] Here is a simplified version of a regulative rule of marriage: If you have publicly committed to a love relationship with X in the presence of an official, then you are obliged to refrain from sexual relationships with any other person Y. As usual, the actual institution encompasses several rules like this at once.

that are usually taken to be radically incommensurable. Philosophers so far have made it rather difficult for social scientists to appreciate the value of their theories. Hubert Knoblauch (1996: 1461) for example has claimed that social scientists can learn from reading Searle's book 'how big the hiatus between philosophy and the social sciences has become', while according to Thomas Osborne Searle's work 'is quite literally *indifferent* as sociology' (1997: 98). In response, Searle could claim that constitutive rule theory and sociological theory are complementary in that the former focuses on social ontology and the latter on explanation. Such a response, however, would be unsatisfactory. Our best guide to ontology is provided by our best scientific theories. According to the widely accepted method of inference to the best explanation, we can infer what exists from the theories that best explain our observations. In light of this, we believe that those doing ontology cannot avoid being concerned with explanation.

A central goal of philosophy is to integrate different perspectives on reality and develop an overall picture. An important avenue along which this goal can be pursued is by building a bridge between the manifest image and the scientific image, to use Wilfred Sellars' terms. And this is exactly what the unified theory purports to do. It tries to connect philosophical theories of property, money, and presidents, to the explanatory insights from the social sciences. The constitutive rule theory illustrates, for instance, how institutional terms influence the way in which we classify objects and persons, as well as the actions that we perform in everyday life. Many of us encounter traffic lights, wedding rings, and signatures on a regular basis. The unified theory reveals how these institutional statuses are related to the regularities in behaviour that feature at the heart of equilibrium theories of institutions. Moreover, it reveals that institutional statuses are conducive to economy of thought.

It is unfortunate that the two approaches have diverged to such an extent that it has impeded communication. In contemporary social ontology the fact that they are tightly connected was lost out of sight. Outlining these connections has been the primary motivation behind this paper, and we very much hope that the unified theory will promote increasing collaboration between philosophers and scientists interested in the ontology of the social world.[28]

References

Anscombe, G.E.M. (1981/1958) 'On Brute Facts', in *The Collected Papers of G.E.M. Anscombe*, Vol. 3, 22-5. Oxford: Blackwell.

Aoki, M. (2001) *Toward a Comparative Institutional Analysis.* Cambridge, Mass.: MIT Press.

Aoki, M. (2007) 'Endogenizing Institutions and Institutional Change', *Journal of Institutional Economics*, 3: 1-31.

Aoki, M. (2011) 'Institutions as Cognitive Media between Strategic Interactions and Individual Beliefs', *Journal of Economic Behavior and Organization*, 79: 20-34.

Aumann, R. (1974) 'Subjectivity and Correlation in Randomized Strategies', *Journal of Mathematical Economics*, 1: 67-96.

Aumann, R. (1987) 'Correlated Equilibrium as an Expression of Bayesian Rationality', *Econometrica*, 55: 1-18.

Bacharach, M. (2006) *Beyond Individual Choice.* Princeton: PUP.

Bardsley, N. (2007) 'On Collective Intentions: Collective Action in Economics and Philosophy', *Synthese*, 157: 141–59.

Belnap, N. (1992) 'On Rigorous Definitions', *Philosophical Studies*, 72: 115-46.

Bicchieri, C. (2006) *The Grammar of Society*. Cambridge: CUP.

Binmore, K. (2005) *Natural Justice.* Oxford: OUP.

Black, M. (1962) 'The Analysis of Rules', in *Models and Metaphors. Studies in Language and Philosophy*, 95-139. Ithaca: Cornell University Press.

Calvert, R. L. (1998) 'Rational Actors, Equilibrium, and Social Institutions', in J. Knight and I. Sened (eds.) *Explaining Social Institutions*, 57-93. Ann Arbor: University of Michigan Press.

Carlsson, H and E. van Damme (1993) 'Global Games and Equilibrium Selection', *Econometrica*, 61: 989-1018.

Conte, A. G. (1988) 'An Essay on Constitutive Rules', *Poznan Studies in the Philosophy of the Sciences and the Humanities*, Vol. 11: 251-7. Amsterdam: Rodopi.

Crawford, S. E. S. and E. Ostrom (1995) 'A Grammar of Institutions', *American Political Science Review*, 89: 582-600.

Demsetz, H. (1967) 'Toward a Theory of Property Rights', *American Economic Review*, 57: 347-59.

Frank, R. (1987) *Passions Within Reason*. New York: Norton.

Gibbard, A. (1990) *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.

Gilbert, M. (1989) *On Social Facts.* Princeton: PUP.

Gintis, H. (2007) 'The Evolution of Private Property', *Journal of Economic Behavior and Organization*, 64: 1-16.

Gintis, A. (2009) *The Bounds of Reason*. Princeton: PUP.

Greif, A. (2006) *Institutions and the Path to the Modern Economy.* Cambridge: CUP.

Greif, A. and C. Kingston (2011) 'Institutions: Rules or Equilibria?' in N. Schofield and G. Caballero (eds.) *Political Economy of Institutions, Democracy and Voting*, 13-43. Berlin: Springer.

Hindriks, F. (2009) 'Constitutive Rules, Language, and Ontology", *Erkenntnis*, 71: 253-75.

Hindriks, F. (2011) 'Language and Society', in I.C. Jarvie and J. Zamora-Bonilla (eds.) *The SAGE Handbook of the Philosophy of Social Sciences*, 137-52. London: SAGE.

Hindriks, F. (2012) 'But Where Is the University?', *Dialectica*, 66: 93-113.

Hindriks, F. (2013) 'The Location Problem in Social Ontology', *Synthese*, 190: 413-37.

Hodgson, G. M. (2006) 'What Are Institutions?', *Journal of Economic Issues*, 15: 1-23.

Knight, J. (1992) *Institutions and Social Conflict.* Cambridge: CUP.

Knoblauch, H. (1996) 'The Construction of Social Reality. By John R. Searle', *American Journal of Sociology*, 101: 1459-61.

Lagerspetz, E. (1995) *The Opposite Mirrors. An Essay on the Conventionalist Theory of Institutions*. Dordrect: Kluwer.

Lewis, D. K. (1969) *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

Lewis, D. K. (1983a) 'Scorekeeping in a Language Game', in *Philosophical Papers*, Vol. I, 233–49. Oxford: OUP.

Lewis, D. (1983b) 'How to Define Theoretical Terms', in *Philosophical Papers*, Vol. I., 78-95. Oxford: OUP.

Mantzavinos, C. (2001) *Individuals, Institutions, and Markets.* Cambridge: CUP.

Marmor, A. (2009) *Social Conventions: From Language to Law*. Princeton: PUP.

Maynard Smith, J. (1982) *Evolution and the Theory of Games*. Cambridge: CUP.

Merton, R. K. (1948) 'The Self-Fulfilling Prophecy', *Antioch Review*, 8: 193-210.

Morin, O. (2011) 'Three Ways of Misunderstanding the Power of Rules', in M. Schmitz, B. Kobow and H.-B. Schmidt (eds.) *The Background of Social Reality.* Dordrecht: Springer, pp. 185-201.

Nichols, S. (2004) *Norms with Feelings*. Oxford: OUP.

North, D. (1990) *Institutions, Institutional Change and Economic Performance.* Cambridge: CUP.

Osborne, T. (1997) 'The Limits of Ontology', *History of the Human Sciences,* 10: 97-102.

Ostrom, E. (1990) *Governing the Commons.* Cambridge: CUP.

Penner, J. E. (1996) 'The "Bundle of Rights" Picture of Property', *UCLA Law Review*, 43: 711-820.

Ransdell, J. (1971) 'Constitutive Rules and Speech-Act Analysis', *Journal of Philosophy*, 68, 385-99.

Rawls, J. (1955) 'Two Concepts of Rules', *Philosophical Review*, 64: 3-32.

Raz, J. (1999) 'Explaining Normativity: On Rationality and the Justification of Reason', *Ratio*, 12: 354-79.

Schelling, T. (1960). *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

Schotter, A. (1981) *The Economic Theory of Social Institutions*. Cambridge: CUP.

Searle, J. R. (1969) *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: CUP.

Searle, J. R. (1995) *The Construction of Social Reality*. New York: The Free Press.

Searle, J. R. (2005) 'What Is an Institution?', *Journal of Institutional Economics*, 1: 1-22.

Searle, J. R. (2010) *Making the Social World.* Oxford: OUP.

Simon, H. (1957) 'Bandwagon and Underdog Effects of Election Predictions', in *Models of Man*, 455–557. New York: Wiley.

Skorupski, J. (2010) *The Domain of Reason*. Oxford: OUP.

Skyrms, B. (1996) *Evolution of the Social Contract.* Cambridge: CUP.

Skyrms, B. (2004) *The Stag Hunt and the Evolution of Social Structure.* Cambridge: CUP.

Skyrms, B. (2010) *Signals*. Oxford: OUP.

Stirling, W.C. (2011) *Theory of Conditional Games*. Cambridge: CUP.

Sugden, R. (1986) *The Economics of Rights, Co-operation and Welfare*, 2nd edition 2004 . Oxford: Blackwell.

Sugden, R. (1998) 'Normative Expectations: The Simultaneous Evolution of Institutions and Norms', in A. Ben-Ner and L. Putterman (eds.) *Economics, Values, and Organization*, 73-100. Cambridge: CUP.

Tuomela, R. (1995) *The Importance of Us*. Stanford: SUP.

Tuomela, R. (2002) *The Philosophy of Social Practices.* Cambridge: CUP.

Tuomela, R. (2007) *The Philosophy of Sociality*. Oxford: OUP.

Turner, S. (2010) *Explaining the Normative*. Cambridge: Polity Press.

Ullmann-Margalit, E. (1977) *The Emergence of Norms.* Oxford: Clarendon Press.

Vanderschraaf, P. (1995) 'Convention as Correlated Equilibrium', *Erkenntnis*, 42: 65-87.

Vanderschraaf, P. (1998) 'Knowledge, Equilibrium and Convention', *Erkenntnis,* 49: 337-69.

Vanderschraaf, P. (2001) *Learning and Coordination.* London: Routledge.

Wagner, E. O. (2012) 'Evolving to Divide the Fruits of Cooperation', *Philosophy of Science*, 79: 81-94.

Young, P. H. (1998) *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton: PUP.

Zollmann, K. J. S. (2008) 'Explaining Fairness in Complex Environments', *Politics, Philosophy and Economics*, 7: 81–98.

University of Milan, Italy, and University of Groningen, The Netherlands.