Journal of Institutional Economics (2015), 11: 3, 459–480
 doi:10.1017/S1744137414000496
First published online 16 October 2014

Institutions, rules, and equilibria: a unified theory*

FRANK HINDRIKS**
Faculty of Philosophy, University of Groningen, Groningen, The Netherlands
FRANCESCO GUALA***
Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Italy

Abstract. We propose a new framework to unify three conceptions of institutions that play a prominent role in the philosophical and scientific literature: the equilibria account, the regulative rules account, and the constitutive rules account. We argue that equilibrium-based and rule-based accounts are individually inadequate, but that jointly they provide a satisfactory conception of institutions as rules-in-equilibrium. In the second part of the paper we show that constitutive rules can be derived from regulative rules via the introduction of theoretical terms. We argue that the constitutive rules theory is reducible to the rules-in equilibrium theory, and that it accounts for the way in which we assign names to social institutions.

1. Introduction

Institutions are ubiquitous. Even a simple description of who we are (two academics) or what we do would be very difficult if we could not use institutional terms such as 'professor', 'university', 'tenure', or 'scientific journal'. Since our behaviour is constantly influenced by institutional entities and institutional roles, institutions have always been a central topic of research in the social sciences. But institutions are also philosophically interesting, for a variety of reasons. Institutions are peculiar products of human activities, to begin with, and may hold the key to understand our special place in the natural world. Why are humans the only animals who can build diverse social organizations and who constantly invent new ways of living together? The other social animals do not seem to have institutions – but then what are we referring to when we talk about institutions? Are they particular patterns of behaviours? Or perhaps

*Parts of this paper have been presented at the University of Helsinki, Erasmus University Rotterdam, Lund University, the University of Turin, the Ecole Normale Superieure in Paris, and at a conference of the Italian Society for Logic and Philosophy of Science. We have benefited from the remarks of many participants, but we are particularly grateful to Mikael Cozic, Conrad Heilmann, Geoffrey Hodgson and two anonymous referees of this journal.
**Email: f.a.hindriks@rug.nl
***Email: francesco.guala@unimi.it

representations of behaviour? Do they have an objective status or are they figments of the human mind? Can they be studied scientifically or do they require some other method of investigation?

Some of these questions are bewildering. So unsurprisingly – in spite of many years of sustained discussion – there is still no agreement on what institutions are. One option of course is simply to ignore these issues and go on studying what institutions do (their function) and how they do it (their mechanics). It is not uncommon for different scientific research programmes to rely on different understandings of key theoretical terms, after all, and such differences may foster healthy competition. At the same time, however, heterogeneous understandings of the basic concepts may hamper communication, making mutual criticism and cross-fertilization difficult.

We think that conceptual heterogeneity is currently an obstacle to communication and collaboration across science and philosophy. A scholar approaching the literature with a fresh mind may have the impression that philosophers and social scientists are talking about completely different things, when they talk about institutions. This impression would be confirmed by the way in which researchers belonging to different camps ignore each other's work, or dismiss it as irrelevant for their own concerns. The most prominent and influential philosophical theory of institutions of the last 20 years, to give an example, has been deemed 'quite literally indifferent as sociology' (Osborne, 1997: 98), while according to another reviewer the same theory shows 'how big the hiatus between philosophy and the social sciences has become' (Knoblauch, 1996: 1461).[1]

We find this state of affair unfortunate, and we believe that it is time to remedy. In this paper, we offer a systematic comparison of the main traditions or conceptions of institutions that inform current research in philosophy and the social sciences. According to the *rule-based* conception, institutions are behavioural rules that guide and constrain behaviour during social interaction, while according to the *equilibrium-based* conception institutions are equilibria of strategic games. The third account of institutions that we shall consider is prominent mostly in philosophy, and conceives of institutions as systems of *constitutive rules* that assign statuses and functions to physical entities – for example pieces of papers that are to be used as money.

We will proceed in two steps. In the first part we will show that the rules and the equilibria approaches are each wanting on their own, but that jointly they provide a satisfactory account of institutions. Following theorists like Masahiko Aoki and Avner Greif, we will combine the best insights of both approaches

1 Osborne and Knoblauch are referring to John Searle's *The Construction of Social Reality* (1995), which will be discussed later in the paper. Searle in turn has repeatedly claimed that social scientists have been unable to address social ontology in a satisfactory manner (see e.g. Searle, 1995: xii; 2005: 1–3; 2010: 200–202).

in a single framework that we call the *rules-in-equilibrium* account. In the second part of the paper we will extend the analysis to the constitutive rules account, showing that it is reducible to the rules-in-equilibrium account. The key step is the demonstration that constitutive rules are nothing but (systems of) regulative rules augmented with the introduction of new theoretical terms. If our argument is correct, it is possible to accomplish the unification of the three main conceptions of institutions within a new theoretical framework. As discussed towards the end of Section 4, the payoff of unification is a significant increase in explanatory power. The history of science demonstrates that 'the explanations of different phenomena most likely to survive are those that can be connected and proved consistent with one another' (Wilson, 1998: 57). Thus, our endeavour is motivated by the idea that an in depth investigation of how theories relate and of whether they can be integrated can have a significant theoretical payoff.

The paper is organized as follows: the rules-based and the equilibrium-based approaches are discussed in Sections 2 and 3. Section 4 describes the rules-in-equilibrium account. The constitutive rules account is analysed in Section 5, and Section 6 explains its relationship with the rules-in-equilibrium conception of institutions. We conclude in Section 7 by explaining what makes a unified theory attractive.

2. The rules account

The most popular and widely cited characterization of social institutions can be found in the opening paragraphs of Douglass North's monograph on *Institutions, Institutional Change and Economic Performance*:

> Institutions are the rules of the game of society or, more formally, the humanly devised constraints that shape human interactions. [. . .] Institutions reduce uncertainty by providing a structure to everyday life. 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)

Like many other scholars, North does not say explicitly whether he is giving a definition, an empirical description, or whether he is introducing an idealized theoretical concept for the study of institutions. Our first goal, therefore, will be to identify and lay bare the 'conception' or 'account' of institutions that is implicit in his work and in the work of other scientists.

The rule-based conception belongs to a venerable tradition that goes back to the founders of modern social science.[2] It states what institutions are (they are

2 The idea of institutions as 'rules of the game' (*Spielregeln*) is already in Weber (1910: 459). On institutions as rules see also Parsons (1935), Knight (1992), Mantzavinos (2001), Hodgson (2006).

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

rules) and what they do (they facilitate human interactions). Consider marriage for instance. Married couples have rights and obligations that indicate what they must and must not do when they engage in certain activities. In most Western countries both husband and wife are responsible for procuring the material resources to support their family, for example. They are both responsible for their kids' welfare and education; they have a mutual right of sexual monopoly, and they are committed to support each other in case of need. The reason why such rules exist is fairly obvious: they help husband and wife attain goals that would be more difficult to accomplish if they acted independently, in an uncoordinated manner.

This idea can be generalized to many other cases: institutional economists like North have used the rules conception to study the way in which institutions facilitate economic growth, for example. Accountancy rules foster transparency and trust; bankruptcy rules reduce uncertainty when businesses fail; property rights encourage investments, and so forth. By inventing and following new rules people can overcome the natural obstacles that limit production, trade, and more generally hinder the welfare of a society. Another virtue of the rules account is that it is closely related to policy. Rules often emerge by trial and error and spread spontaneously by imitation, but they can also be designed and implemented by an authority by issuing laws and decrees.

Many rules, however, are never followed even though they are formally included in the legal system. In May 2010, for instance, ten French ministers proposed to repeal a law that forbids women to wear trousers. The law had been in place since 1799, although no one had tried to implement it for a long time. Rules like the French ban on trousers are *ineffective*, and raise a difficult problem for rule-based accounts of institutions. Why are some rules followed, while others are not?

In the case of the French ban on trousers, the law was simply forgotten. But some formal rules such as the speed limit are rarely observed even though they are universally known. In many North-American states many cars drive between 65 and 75 mph, for example, in spite of an official speed limit of 65 mph (Greif and Kingston, 2011). So clearly the formal rule is not effective – the real, informal speed limit is somewhere around 75 mph. But to say that 65 mph is not the 'real' rule leaves several important questions unanswered: What distinguishes 'real' from merely 'nominal' rules? What is the difference between the 65 mph rule and the 75 mph rule? Why do people comply with the latter but not with the former?

A sketchy explanation may go like this: the nominal rule is just a signal that indicates roughly what kinds of behaviours are expected, but no one believes that it will be followed strictly. It would be pointless for the police to fine all the drivers who exceed the official limit by a small margin (those who drive at 67 mph, say). It may be wise to sanction only major violations of the nominal rule and implement a stochastic strategy: fine every car speeding at 75 mph or

more; fine some cars speeding around 75 mph; fine no car speeding at 65–70 mph. This would work reasonably well and would ensure that most people do not exceed 75 mph. Drivers have an incentive not to exceed 75 mph; the police has an incentive to tolerate those who do not exceed the 75 mph limit. If a naïve observer were to look at the traffic flowing down the highway, she would conclude that the effective speed limit is roughly 75 mph: everybody believes that one should not exceed that limit, and everybody's behaviour confirms that belief. The system is in equilibrium.

The preceding line of argument puts some pressure on the idea that institutions are rules. It suggests that they are perhaps better conceived of as actions that people have an incentive to make. As a consequence, one might think that an account based on the concept of equilibrium can incorporate incentives and make rules redundant. Rules cannot be institutions, the thought would be, because by themselves they lack the power to influence behaviour.[3]

3. The equilibria account

The idea that institutions are equilibria of strategic games is central to another account of institutions that has been prominent in the literature for the last three decades.[4] 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 non-cooperative 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.

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 by a simple sentence of the form 'choose X' or 'do Y'. The defining characteristic of an equilibrium – what distinguishes it from other profiles – is that each strategy must be a best response to the actions of the other players or, in other words, that no player has an incentive to change her strategy unilaterally. If the others do their part in the equilibrium, no player can do better by deviating. Those who defend a pure equilibria account hold that institutions can be equated with equilibria that have certain properties. They maintain that recourse to rules is not needed. We will argue that this will not do, because rules play an essential role in achieving those equilibria that form institutions.

3 Notice that even though the rules account has difficulty explaining widespread deviation, it succeeds in capturing the fact that the *codified* speed limit is 65 mph. By putting symbolic codification at center stage, the rules account captures an important aspect of institutional reality that must be retained by any theory that attempts to supersede it. We shall return to this point in Section 4.

4 See e.g. Lewis (1969), Ullmann-Margalit (1977), Sugden (1986), Skyrms (1996, 2004), Calvert (1998), Young (1998), Aoki (2001), Vanderschraaf (2001), Binmore (2005), Bicchieri (2006).

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

Figure 1. The private property game (hawk-dove).

	U	NU
U	0, 0	2, 1
NU	1, 2	1, 1

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 simple model for the institution of private property.

The use of resources such as land raises a coordination problem if interests are served badly by two persons trying to use the same piece of land. The optimal solution in such cases is that one uses the land, perhaps to grow a crop, and the other abstains from using it to graze her cattle. The game of private property can be represented in strategic form using a matrix known as 'hawk-dove' in biology, and 'chicken' in economics.[5] For every piece of land, the players have to make a decision: in Figure 1 the strategy U stands for 'use', NU for 'not use'. If they both decide to use the same land, the players will end up fighting, which is the worst outcome for all (0, 0).[6] If they both abstain, they will not clash but will miss the opportunity to use the land (1, 1). The best solution is to converge on one of the two equilibria in the top-right and bottom-left corners, where one player uses the land and the other lets him use it.

The property 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.[7]

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

6 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).

7 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

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

An obvious solution in such circumstances is to adopt a *correlation device*. A correlation device is a signalling mechanism that the players can use to coordinate their actions. Think of a traffic light, for example, indicating by means of different colours (red/green) who has the right to cross a busy road at each particular moment. In general, correlation devices need not be artificial tools built for a specific purpose. Any external mechanism may do, provided it sends reliable and correlated signals to the players. In the case of property, for example, the players may rely on a simple pre-emption device: whoever occupied the land first has the right to use it. The temporal order of occupation, or the sequence of the claims made by the players, is used as correlation device. Except in rare cases, this device provides unambiguous, correlated signals to the players. If they all follow this simple mechanism, fights should be avoided and coordination should run smoothly. No agent is served better by acting differently, on the assumption that the others follow the signal, which implies that the set of actions is a *correlated equilibrium*.[8]

Technically speaking, this solution involves a set of *conditional* strategies. Each player conditions her move (U or NU) on her temporal and physical location relative to the piece of land. If she arrived first, the player uses the land, if she did not, she lets the other player use it. In a series of repeated encounters, the average payoff the players achieve will depend on the probability of occupying the land first. If the probability is roughly equal, for example, they will both achieve an average payoff of 1.5 units in the long run. But even if the outcome does not respect perfect equality, the correlated equilibrium of the property game tends to be more egalitarian than either of the two asymmetric equilibria of the hawk-dove game with uncorrelated (or unconditional) strategies.

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. The only difference is that in the driving game the conditional strategy does not lead to a substantially different outcome than any of the two unconditional strategies ('keep right', 'keep left'). In the property 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.

We will assume, for the sake of the argument, that this story gives an adequate (albeit simplified) account of the institution of property. Since *real* property rights involve more than the right to use, we shall use a star symbol to distinguish our

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 U or NU with probability 1/2. This equilibrium delivers expected payoffs of one unit each and is, therefore, inefficient. We will ignore the mixed-strategy equilibrium from now on.

8 See Aumann (1974, 1987), Vanderschraaf (1995, 1998, 2001), and Gintis (2007, 2009).

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

Figure 2. A prisoner's dilemma and a stag hunt game.

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

Prisoner's Dilemma

	S	H
S	2, 2	0, 1
H	1, 0	1, 1

Stag Hunt

simple proto-institution from its real-world counterpart. Property* is a correlated equilibrium of the hawk-dove game, in which one player uses and the other one refrains from using a piece of land, according to the pre-emption system.

Not all equilibria can be institutions, however. Two features of the private property game are important: first, it is a coordination problem; and second, the solution requires that the players correlate their strategies. The significance of coordination can be illustrated using a prisoner's dilemma game (Figure 2 on the left). Consider mutual defection. The pair of strategies (D, D) is an equilibrium, but intuitively it is not an institution. Why? The reason is that each agent can implement the rule independently. There is no need to coordinate strategies. In fact there is no reason to even think about the action of the other player: whatever she does, it is optimal to defect. That's why mutual defection in this game is often taken to represent the proto-typical failure of sociality.

To appreciate the significance of correlation, consider the stag hunt game on the right of Figure 2: both (S, S) and (H, H) are equilibria of this game. But (H, H) does not require that the players correlate their strategies. The minimum payoff is guaranteed, so one does not have to pay attention to what the other player does. Since in (S, S) correlation is crucial, but in (H, H) it is not, only the former equilibrium is an institution.

4. The rules-in-equilibrium account

So institutions must be *correlated* equilibria of *coordination* games with multiple equilibria. In a correlated equilibrium, as we have seen, the strategy of each player is conditioned on an event or signal sent by a coordination device. In order to achieve a satisfactory definition of institutions, however, we must introduce a third condition: *representation*, to capture the idea that the players must be able to represent the equilibrium in symbolic form. As we discuss in Section 5, this can be facilitated by so-called constitutive rules that have special symbolic significance. The reason why this third condition of representation is needed is that the notion of correlated equilibrium is far too permissive and would let in too many behavioural patterns that we would not intuitively consider institutions.

So-called animal conventions are a paradigmatic example. Consider *Pararge aegeria*, a butterfly living in the woodlands of Asia and Europe. Male butterflies patrol the patches of sunlight that appear on the woodland's floor, where they

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

mate with females after a brief courtship. When a male enters a sunspot that is already occupied by another male, the incumbent attacks it. After a brief skirmish, the defeated butterfly leaves the spot. Remarkably, the intruder is nearly always defeated, and the incumbent nearly always retains its territory.[9]

Pararge aegeria play correlated equilibria, and similar behavioural patterns have been observed in swallowtails, baboons, and lions. The standard interpretation is in terms of a repeated hawk-dove game. As a solution, these species have evolved pairs of strategies ('conventions') that minimize damage by granting the territory and the mating opportunity to the incumbent after a ritual contest. The biologist John Maynard-Smith (1982), who first used game theory to explain such behaviour, has called it 'bourgeois equilibrium'.

Non-human animals solve coordination games using correlation devices, but animals do not have institutions. Since there are in nature a few examples of anti-bourgeois equilibria, the expression 'animal convention' seems to be more appropriate. Whichever equilibrium has been selected, however, the important point is that the strategies are biologically implemented or – in a broad sense of the term – genetically encoded in each species. A group of butterflies cannot coordinate on anything but who occupied the sunspot first. They can play *this* particular strategy only. They cannot invent a new equilibrium. Humans, in contrast, can: they hook onto different correlations, invent constantly new strategies, and dramatically enlarge the number of possible equilibria.

What distinguishes human institutions from the correlated equilibria of *Pararge aegeria*? The answer seems to be that butterflies react only to a narrow set of signals, such as who enters the spotlight first. A simple mechanism that links one type of stimulus with one type of behaviour guarantees coordination. More complex creatures in contrast are able to decouple stimulus and behaviour. They do so by adding an intermediate state – a *representation* of the environment – that they use to condition their behaviour (Sterelny, 2003). Moreover, such complex creatures can condition their strategies on *many* different representations – many signals and many correlation devices. In the case of humans, we say that they can follow different *rules*. These rules are representations in symbolic form of the strategies that ought to be followed in a given game. Just like the rules account without equilibria is incomplete, so is an equilibrium-based account without rules. A satisfactory theory must combine the best features of both.

Notice that the concept of a rule is ambiguous. Sometimes we use rules to describe, and sometimes to prescribe behaviour; occasionally we use them to do both things at once. But these functions are conceptually distinct. Let us distinguish between agent-rules (or a-rules for short) and observer-rules (o-rules), respectively. An observer formulates an o-rule mainly to summarize others' behaviour; an agent formulates an a-rule to summarize and to guide her own

9 The classic study is Davies (1978).

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

behaviour.[10] Equilibrium theories are observer theories, and so the actions of the players are described from an observer's point of view only. Formulating rules, however, may also facilitate convergence on an equilibrium, so an adequate theory must capture the fact that rules are used both to represent and to influence behaviour.

These insights can be combined into a coherent whole by stipulating that institutions are *rules in equilibrium*, where the rules are summarized by the agents using some kind of symbolic representation. According to Avner Greif and Christopher Kingston, for example

> Despite their differences, the institutions-as-rules and institutions-as-equilibria approaches have much in common and are best viewed as complements rather than substitutes. (2011: 15)

> [...] the role of 'rules', like that of other social constructs, is to coordinate behavior. Because there are multiple potentially self-enforcing expectations in a given situation, coordination mechanisms, including rules, play an essential role in generating regularities of behavior and social order. Rules fulfill this coordinating role by specifying patterns of expected behavior, and also by defining the cognitive categories – signs, symbols, and concepts – on which people condition their behavior. (2011: 28)

In a similar vein, Masahiko Aoki (2007, 2011) emphasizes the importance of public representations or social cognitive artefacts. He proposes the following definition:

> An institution is a self-sustaining, salient pattern of social interaction, as represented by meaningful rules that every agent knows, and incorporated as agents' shared beliefs about the ways the game is to be played. (Aoki 2007: 6)

In a nutshell, the rules represent equilibria (or parts of equilibria) and help the players to exploit a particular correlation device. Let us see how this account works in the simple case of property*. Recall that the players (P1 and P2) use pre-emption as a correlation device. The correlated equilibrium in the game of property* is the pair of strategies:

(s_1) Use if P1 occupied first, do not use if P2 occupied first.

(s_2) Use if P2 occupied first, do not use if P1 occupied first.

From the point of view of an external observer, the convention that regulates property* takes the form of a regularity that corresponds to a correlated

10 Knight (1992: 69) makes a similar distinction between 'regularities' and 'rules', where the former are said to be essentially backward-looking and the latter forward-looking. Ostrom (1990: 51) and Hodgson (2006) also emphasize the guiding role of institutions, and offer analyses that are in many ways compatible with our 'rules-in-equilibrium' account. For a seminal discussion of different kinds of rules, see Rawls (1955).

equilibrium in the hawk-dove game. But each strategy in this profile also takes the form of a rule that dictates each player what to do in the given circumstances. Each player, therefore, will perceive the institution as a prescription to use the land if the circumstances are 'right'. Since the two strategies are formulated as rules, clearly the equilibrium is a set of rules – one for each player – that, as North puts it, 'establish a stable structure to human interaction' (1990: 6).

Unlike in 'pure' rules-based theories, the concept of equilibrium is central in this account. But unlike 'pure' equilibrium-based theories, this account brings at centre stage the representation of the equilibrium strategies by means of symbolic markers (rules). Aoki (2001, 2007) in particular has emphasized that symbolic markers summarize the properties of equilibria. Institutions help individual players not only to reach coordination, but also to economize cognitive effort. As we shall see shortly, one way of doing this is simply by means of theoretical terms that are used to encompass an entire class of rules under the umbrella of a single institution. This process – the naming or baptizing of institutions – has been analysed in depth by philosophers and will be discussed in detail in the remaining sections of this paper.

Before doing so, let us pause briefly and comment on what has been achieved thus far. We have argued that both the equilibrium approach and the rules approach can capture certain aspects of institutions, but that a proper account requires a combination of both. Furthermore, we have proposed a unified theory that does indeed combine both. The fact that this can be done reveals that they are not inconsistent, but complementary. We have shown that existing theories fail to do justice to the role of either coordination, or correlation, or representation. The unified theory, in contrast, provides an adequate explanation of institutions, because it incorporates all three dimensions.

As it combines insights from both approaches, the explanatory power of the unified theory is larger than that of theories that belong to either one of them. One advantage is explanatory efficiency: the theory explains more aspects of institutions than its rivals. However, explanatory power is not only a matter of convenience. It also serves to reveal that apparent diversity can be traded for an appreciation of the actual unity of the social world (Mäki 2001: 502–03). This can be done by increasing explanatory depth or explanatory integration (Ylikoski and Kuorikoski 2012). Providing a mechanism, as we have done when discussing correlation, increases explanatory depth. The unified theory can answer more questions than the two original approaches could do even in combination. As a consequence, it provides a higher degree of understanding of the nature of institutions.

We have seen that some theorists have already tried to integrate aspects from the rules and equilibria approaches, and we have followed their lead until now. But in the next few sections we will take a crucial step forward: we will incorporate a third approach that focuses on the representational or symbolic dimension of institutions, and which has become increasingly influential during

the last two decades, especially in philosophical circles. Although the approach is a variant of the rule-based account of institutions, it attempts to explicate institutions using a very different kind of rule that, instead of merely regulating behaviour, creates the very possibility of new types of behaviour. Our goal in the remaining part of this paper is to demonstrate that this approach – the *constitutive rules* approach – can be encompassed within the theory of institutions as rules-in-equilibria.[11]

5. The constitutive rules account

The best-known proponent of the constitutive rules approach 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). Searle contrasts constitutive rules to *regulative rules* that have as their syntax 'do X', or 'if Y, do X'. Since the actions or strategies that appear in game-theoretic accounts of institutions have precisely this form, regulative rules play a key role in the rules-in-equilibrium approach. So Searle's distinction is meant to suggest that there is a deep hiatus between his approach and the accounts of institutions found in the social science literature.

A central claim of the constitutive rules approach is that institutions exist only because we believe they exist. Our beliefs are thought to play a constitutive role with respect to institutional actions. The constitutive view, however, is not restricted to actions. In addition to actions, institutions often involve objects (like money, university buildings), persons (police officers, presidents), and events (declarations of war, graduations). The constitutive view applies to items from all of these ontological categories. In the case of money, for example, a constitutive rule is: 'Bills issued by the Bureau of Engraving and Printing (X) count as money (Y) in the United States (C)' (1995: 28).[12] The schematic letter X can be replaced by predicates that apply either to actions or to items from several other ontological categories. But what does the letter Y refer to, exactly?

Money, according to Searle, is an example of a *status function*. By accepting certain entities as money we assign the status function of being a means of exchange to these entities. For the purposes of our analysis, it will be useful to break Searle's formula in two parts, introducing the twin notions of 'base rule' and 'status rule'. A *status rule* explicates what it means to have a certain status. The status rule of money, for instance, is 'money is a medium of exchange'. A *base rule* specifies what it takes to have a certain status. In certain contexts, an

11 Hindriks (2005, chapter 7) first argued that the constitutive rule theory and Lewis' equilibrium theory can be unified with one another. In this paper, we build on that argument and extend it to a wider range of theories about institutions.

12 Such bills are in fact issued by the Federal Reserve.

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

item has to be a shell in order to be money; in others it has to be a disc of metal, and so forth. In more theoretical terms, base rules concern the ontological basis or constitution base of statuses. Status rules, in contrast, are meaning rules and concern the definition of status terms. They concern the behaviour that the status regulates, or the rights and obligations that the status entails.

The 'counts as' phrase that appears in Searle's formulation of a constitutive rule can now be interpreted more precisely by relating it to base rules. In medieval Finland, for example, squirrel pelts were money (Tuomela, 2002). Searle would say that the constitutive rule of money in medieval Finland was 'Squirrel pelts count as money in Finland'. We suggest interpreting the 'counts as' phrase as follows:

$$\text{X counts as Y} \leftrightarrow \text{X is collectively accepted as Y}$$

and

$$\text{X is collectively accepted as Y} \leftrightarrow \text{X is Y.}$$

In our terminology, the base rule of money relevant to medieval Finland takes the following form: 'In Finland, squirrel pelts are money'. This base rule applied in the Middle Ages because it was collectively accepted to apply.

The notion of collective acceptance has been proposed in relation to that of collective intentionality – roughly, the intentionality exhibited by social groups.[13] Standard game-theoretic approaches do not deploy such a notion. However, collective acceptance can be dissociated from the notion of collective intentionality and defined in general terms as whatever set of intentional states is needed for institutions. Thus, the standard game-theoretic notions of preferences, expectations and common knowledge may qualify as a kind of collective acceptance.[14]

With this proviso in mind, let us address the distinction between regulative and constitutive rules. For the sake of concreteness, it is convenient to focus on a specific example, so we shall return to our proto-institution of property*. In Section 4 we analysed this institution by means of the correlated equilibrium (or pair of strategies):

(s_1) Use if P1 occupied first, do not use if P2 occupied first;

13 There are many theories of collective intentionality in the literature, see, for example, Gilbert (1989), Searle (1990, 2010), Bratman (1992, 1993), Tuomela (1995, 2002).

14 In connection to the example of money, in fact, Lewis (1969: 49) uses the term 'accept' himself. On the relationship between game-theoretic notions of collective beliefs and philosophical theories of collective intentions, see e.g. Bacharach (2006), Bardsley (2007), Gold and Sugden (2007), Hakli *et al.* (2011). The reducibility of collective to individual intentions is a thorny issue in the philosophy of action, for according to some authors preference, beliefs and common knowledge conditions do not do justice to the normative dimension of institutions (see e.g. Tuomela 2002: 128–29). At the same time, however, such issues are orthogonal to the main topic of this paper. It is perfectly possible to discuss the relation between the rules-in-equilibrium and the constitutive rules approach while remaining neutral on this matter.

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

(s_2) Use if P2 occupied first, do not use if P1 occupied first,

where the players use the pre-emption system to coordinate their actions. We have remarked earlier that these two strategies appear to the relevant players as rules that guide and constrain their actions in the game of private property (hawk-dove). These rules are regulative rules, in Searle's language, and for simplicity they can be summarized by means of a single principle:

[R] If one is the first to occupy a piece of land, one has the right to its exclusive use.

Notice that this rule does not include a label for or name of the institution. Suppose we now introduce the term 'property*' as follows: we say that what it takes for a piece of land to become someone's property* is that she is the first to occupy it. Furthermore, we say that what it is or means for a piece of land to be someone's property* is that she has the right to its exclusive use. By so doing we have split the regulative rule in two parts and used the term 'property*' to turn these parts into complete sentences: the first one says that a piece of land is the property* of the person who is the first to occupy it; the second one that if a piece of land is someone's property*, she has the right to its exclusive use.

Another way to put it is that we have transformed the regulative rule [R] in two rules, [B] and [S], respectively:

[B] If a person first occupies a piece of land, then it is her property*.

[S] If a piece of land is someone's property*, she has the right to use it.

Now, the combination of these two rules forms a constitutive rule:

[C] If a person first occupies a piece of land then it is her property*, and if a piece of land is someone's property* then she has the right to use it.

The [C] rule has the typical structure of Searle's 'X counts as Y in C' formula, provided that (a) the expression 'counts as' is interpreted in terms of conditions of acceptance, as proposed earlier; (b) the Y term is unpacked so as to make the content of the status function explicit by means of the status rule. A constitutive rule, once these two points have been made explicit, has the following structure: 'If C then X is Y, and if Y then Z', where 'if Y then Z' is a status rule that specifies the actions that are made available to the relevant individuals. The view that regulative rules can be transformed into constitutive rules using this XYZ schema and via the introduction of terms such as 'property*' is what we call (following Hindriks 2005, 2009) *the transformation view of constitutive rules*.

The same procedure can be used to introduce other terms, referring, for example, to institutional roles. We may create a rule stating that 'The person who is the first to occupy a piece of land owns* it' and another one stating that 'An owner* has the right to exclusive use of her land', for example. Transforming a regulative rule by introducing institutional terms such as 'owner*' or 'property*'

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

is very convenient. On the supposition that the stipulated usage of the term is generally accepted, the simple claim that a particular piece of land is my property* conveys a lot of information. It presupposes that I was the first to occupy it and it means that I have the right to its exclusive use. Thus, in line with the rules-in-equilibrium approach, the representation of the equilibrium in symbolic form has the advantage of cognitive economy, especially in those cases where several rules are used to govern interrelated strategic interactions. But apart from this, no big changes are implied as far as the original rule is concerned. In particular, behaviour in accordance to the original rule [R] is extensionally equivalent to behaviour in accordance to the content of the rules [B] and [S] that employ the term 'property*'.

6. Transformation, elimination, and the reference of theoretical terms

The argument presented in the previous section, if correct, entails that the rules-in-equilibrium and the constitutive rules approaches are perfectly consistent. Constitutive rules are linguistic transformations of regulative rules. Such transformations rely on the introduction of a new term that is used to name an institution. In the end, constitutive rules are nothing but (systems of) regulative rules augmented by the introduction of theoretical terms.

In this section we address a worry one might have about the transformation view, and we review a number of virtues of the rules-in-equilibrium approach. To begin with the former, it may be argued that some transformations introduce qualitative changes that preclude consistency between the rules before and after the transformation. To understand this worry, let us draw an analogy with theoretical revolutions in science: a paradigm shift is generated sometimes by introducing new theoretical terms and abandoning some terms that played a key role in an old scientific theory (Kuhn 1970). The introduction of new terms may change the meaning of the original terms that survive the transformation, and as a consequence the post-transformation theory may be inconsistent with the pre-transformation theory.[15] So, we need to find transformation criteria that guarantee consistency.

Belnap's (1993) criteria for rigorous definitions can play this role. 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 favour 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 it does not entail anything

15 For an intuitive example, consider how the meaning of 'weight' was changed by the introduction of the term 'mass' in Newtonian physics.

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

new – anything that is qualitatively different from what can be expressed by only using terms previously understood. We will say that the addition of a term is a *conservative transformation* of the theory in which it is used, if the definition of that term is rigorous in this sense.[16]

The core of the transformation in the case of institutions is the introduction of a Y-term. This introduction as we have seen leaves all the features of the rules-in-equilibrium account intact. In other words, the transformation does not introduce any alterations that are in conflict with the theory that explicates institutions using only rules in equilibrium (strategies). Another way to put it is that constitutive rules are regulative rules with special features. In particular, they are regulative rules that have been split in two parts using Y-terms to turn the parts into complete sentences. And the definitions of Y-terms are rigorous in Belnap's sense: they do not lead to anything that could not be obtained before. Before the introduction of the relevant Y-term the link between the two parts was internal to the regulative rule. After its introduction, the link is forged by the fact that the Y-term figures both in the base rule and in the status rule. This implies that definitions of Y-terms are conservative in Belnap's sense.

Y-terms are also eliminable. A constitutive rule can be transformed into a regulative rule by reversing the transformation process outlined above. The first step is to eliminate the Y-term, and the second to join the resulting parts to form a complete sentence. In other words, the thing to do is to move back from [C] to [B] and [S], and from these to [R]. Thus, the definition of a Y-term also satisfies the criterion of eliminability. This implies that at the level of reference there are no substantial changes: the behaviour implied by [R] is extensionally equivalent to that involved in following [B] and [S]. Given that nothing that conflicts with the rules-in-equilibrium account has been introduced, we can conclude that the constitutive rules approach and the rules-in-equilibrium approach are perfectly consistent.

As conservative transformations do not involve qualitative changes of the theory at issue, why should we bother to introduce the theoretical terms in the first place? In other words, one might worry that this argument shows too much: if the rules-in-equilibrium and the constitutive rules accounts are in a sense equivalent to one another, one may conclude that the constitutive rules account has nothing to offer that the rules-in-equilibrium account does not have. This conclusion, however, would be hasty. Constitutive rules are useful theoretical constructs that help us understand several important features of institutions. In the second half of this section, we highlight four virtues of the rules-in-equilibrium approach that are closely associated to the use of theoretical terms. We focus in particular on four virtues, two of which are linguistic and two ontological. These virtues exhibit the explanatory power of our unified theory

16 If it is – and the theory is adequate in other respects - the theoretical term is bound to refer (Lewis, 1983). See the end of this section for more on this.

https://doi.org/10.1017/S1744137414000496 Published online by Cambridge University Press

of institutions: they reveal some more of the advantages of the unification that we achieve in this paper.

Language: analysis and representation

An important reason for taking the transformations of regulative rules seriously is that we do in fact employ many Y-terms. 'Money' and 'marriage' as well as 'president', 'property', and 'promise' are prominent examples. The version of the constitutive view under discussion helps us to appreciate the meaning of these terms and analyse the way they are used in particular contexts. By investigating such terms using the XYZ-schema we can do justice to the linguistic framework with which ordinary people operate in their everyday lives. Another way to put it is that the transformation view builds a bridge between the ontology implicit in ordinary language and the ontology of social science. And the constitutive rules account plays an important role in this process of unification, by outlining the fundamental grammatical form of the ordinary sentences that contain institutional terms.

Institutional terms are not only interesting for those philosophers who are devoted to the analysis of ordinary language. They are also important to scientists and philosophers interested in explanation. As we have already mentioned, an important function of institutions is to promote economy of thought. Status terms summarize in compact form sets of strategies that would otherwise require considerable cognitive effort. How people achieve this and what gain it offers requires explanation. In this spirit, Aoki points out that institutions are symbolic representations of salient equilibria. And one obvious way to represent equilibrium strategies symbolically is simply to give them a name. As discussed in Section 3, representation distinguishes human institutions from animal conventions, and the transformation view elucidates the way in which people baptize equilibria by means of institutional terms.

Ontology: multiple realizability and parsimony

The unified theory also enables us to see the connections between multiple realizability as discussed by philosophers and multiple equilibria in game theory. Multiple realizability is a much-discussed phenomenon in metaphysics, the philosophy of mind, and the philosophy of the social sciences (Fodor, 1974; Sawyer, 2002). A multiply realizable property is a property that can occur in different guises depending on the context. Money, for example, can come in the form of shells or coins. Similarly, different countries have different requirements for getting married (think of age requirements, for example). In the case of property, the way in which the land is divided is variable, and the criteria of ownership may also vary according to custom or legislation.

The constitutive rules account accommodates multiple realizability by allowing for different base rules for different contexts: one base rule might pertain to shells, while another pertains to coins. Multiple realizability can also

occur within one and the same context, in which case the X-term should be specified disjunctively: a base rule might mention both coins and pieces of paper with certain characteristics as bases for money. Thus, the constitutive view can do justice to the multiple realizability of institutional properties.

What is striking about this phenomenon is its intimate relation to the existence of multiple equilibria in game theory. The multiple equilibria in the case of the institution of property*, for example, correspond to the different ways of dividing up the land. Similarly for many other institutions. Think for example of different items that are used as money in various contexts and of different procedures one has to go through in order to get married in various countries. Thus, unification of the constitutive view and the rules-in-equilibrium account allows seeing that in the case of institutions the ontological phenomenon of multiple realizability is intimately related to the theoretical phenomenon of multiple equilibria. At least in many cases, they come down to the same thing.

There is a sense, thus, in which the rules-in-equilibrium account is parsimonious. Contrary to Searle's repeated remarks, it is not true that the social science accounts based on equilibria and regulative rules do not have the means to distinguish institutional from non-institutional behaviour. Searle is committed to this view because he believes that constitutive rules are not only sufficient for creating (the possibility of) institutional forms of behaviour, but also that they are necessary. The twin claim is that regulative rules are insufficient.

Searle supports this claim primarily by examples. He takes it to be obvious that the rules of etiquette are regulative rules, whereas the rules of chess are constitutive rules. Although these examples have strong intuitive force, they also pose a problem: identifying institutions with constitutive rules in contradistinction to regulative rules implies that Searle must deny that practices of etiquette are institutions. There are no good grounds for believing this to be true. The alternative option of allowing for institutions to consist of regulative rules instead is much more attractive. This is an example of how unification can serve to reveal the underlying unity of the world. The distinction between regulative and constitutive rules obscures the fact that both etiquette and chess are institutional phenomena. A better appreciation of the relation between regulative and constitutive rules makes clear that they are cogwheels of the same social machine, even though they display different grammatical forms.

Collectively accepting a constitutive rule is sufficient for creating (the possibility of) new forms of institutional behaviour. Thus, there is a sense in which constitutive rules can have ontological import. This sense is limited, however, because the same thing could be achieved by collectively accepting a regulative rule. The difference would be a difference in description. Given that [S] and [B] are in force, we have a special term for the institution concerned, 'property*'. The acceptance of [B], which implies the applicability of [S], has ontological import in that it is constitutive of the institution of property*. However, the same institution would exist had [R] been accepted instead.

Contrast this with driving on the right hand side of the road. This rule is part of the rules of traffic. We do not have a special term for the status involved in it, however. Hence, we are limited to describing this institution in terms of a regulative rule, the institution that requires us to drive on the right hand side of the road (the alternative would be to invent a term for it).

David Lewis' views on ontology and scientific realism are congenial to most of what has been said thus far. Lewis would agree that institutions such as money, property, and driving on the right are real phenomena. As a pioneer of the equilibrium approach, he would agree that accepting regulative rules (strategies) suffices for their existence. In addition to this, however, he would also accept that such institutions could be adequately described in terms of constitutive rules. Lewis' work on theoretical terms offers the conceptual resources for this claim: a set of constitutive rules that pertain to a set of interlocking institutions can be regarded as a theory of sorts. If the constitutive rule for property* is part of the set, then 'property*' is a theoretical term. According to Lewis, 'If [a theoretical term] purports to name something, then if the theory that introduced it is true it does name something' (1983: 79). Now, all that constitutive rules do in comparison to regulative rules is to introduce labels or names (such as 'money' or 'property') for the statuses that figure in those regulative rules. Just as physicists could do without theoretical terms, we could do without institutional terms. This does not mean that these terms do not refer: as long as the relevant theories are true, they do name something. All this implies that, if we have formulated the constitutive rules that make up the theory correctly – i.e. if we have transformed the regulative rules into constitutive rules appropriately – then 'property' names something, which means that the institution of private property exists.

7. Conclusion

Our unified theory of institutions aims at encompassing and preserving the best insights of three approaches that have dominated debates in the social science and philosophy literature. In the first part of this paper we have merged the rule-based and the equilibrium-based approaches, along the lines of Aoki (2001, 2007, 2011) and Grief and Kingston (2011), into a unitary framework that we have called the 'rules-in-equilibrium' account. In the second part of the paper we have shown that this framework is perfectly compatible with the approach based on constitutive rules proposed by Searle (1995, 2010). The crux of the argument is that constitutive rules can be created at will from more fundamental building blocks – regulative rules or game-theoretic strategies – via the introduction of institutional terms. This way, we have achieved a unified theory of institutions.

There are two reasons for taking seriously the project of integrating the scientific and philosophical perspectives as suggested in this paper. The first reason is simply the intrinsic value of attaining a comprehensive view of the social world. Such a view can improve our understanding 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 our common-sense ontology of money, property, and presidents to the explanatory insights from the social sciences.

Second, our unified theory makes it possible to appreciate the commonalities and differences between theories 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. As a consequence, their views have been ignored or dismissed as irrelevant by scientists, as we have seen. One possible reaction is simply to retrench and claim that philosophical theories are complementary to social science in that the former focus 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. The unified theory that we propose in this paper explains not only what ordinary people talk about when they talk about institutions, but also how these objects relate to the theoretical constructs that social scientists use to explain coordinated behaviour in complex strategic interactions. We very much hope that this framework will promote increasing collaboration between philosophers and scientists interested in the ontology of the social world.

References

Aoki, M. (2001), *Toward a Comparative Institutional Analysis*, Cambridge, MA: 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, NJ: Princeton University Press.

Bardsley, N. (2007), 'On Collective Intentions: Collective Action in Economics and Philosophy', *Synthese*, **157**: 141–159.

Belnap, N. (1993), 'On Rigorous Definitions', *Philosophical Studies*, **72**: 115–146.

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

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

Bratman, M. (1992), 'Shared Cooperative Activity', *Philosophical Review*, **101**: 327–341.

Bratman, M. (1993), 'Shared Intention', *Ethics*, **104**: 97–113.

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

Davies, N. B. (1978), 'Territorial Defence in the Speckled Wood Butterfly ('Pararge Aegeria'): The Resident Always Wins', *Animal Behavior*, **26**: 138–147.
Fodor, J. A. (1974), 'Special Sciences (Or: the Disunity of Science as a Working Hypothesis)', *Synthese*, **28**: 97–115.
Gilbert, M. (1989), *On Social Facts*, Princeton, NJ: Princeton University Press.
Gintis, A. (2009), *The Bounds of Reason*, Princeton, NJ: Princeton University Press.
Gintis, H. (2007), 'The Evolution of Private Property', *Journal of Economic Behavior and Organization*, **64**: 1–16.
Gold, N. and R. Sugden (2007), 'Collective Intentions and Team Agency', *Journal of Philosophy*, **104**: 109–137.
Greif, A. and C. Kingston (2011), 'Institutions: Rules or Equilibria?', in N. Schofield and G. Caballero (eds.), *Political Economy of Institutions, Democracy and Voting*, Berlin: Springer, pp. 13–43.
Hakli, R., K. Miller and R. Tuomela (2011), 'Two Kinds of We-Reasoning', *Economics and Philosophy*, **26**: 291–320.
Hindriks, F. (2005), *Rules and Institutions: Essays on Meaning, Speech Acts and Social Ontology*. PhD Dissertation, Erasmus University Rotterdam.
Hindriks, F. (2009), 'Constitutive Rules, Language, and Ontology', *Erkenntnis*, **71**: 253–275.
Hodgson, G. M. (2006), 'What Are Institutions?', *Journal of Economic Issues*, **15**: 1–23.
Knight, J. (1992), *Institutions and Social Conflict*, Cambridge: Cambridge University Press.
Knoblauch, H. (1996), 'The Construction of Social Reality. By John R. Searle', *American Journal of Sociology*, **101**: 1459–1461.
Kuhn, T. S. (1970), *The Structure of Scientific Revolutions*, 2nd edn., Chicago, IL, University of Chicago Press.
Lewis, D. (1983), 'How to Define Theoretical Terms', in *Philosophical Papers*, Volume I, Oxford: Oxford University Press, pp. 78–95.
Lewis, D. K. (1969), *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.
Mäki, U. (2001), 'Explanatory Unification: Double and Doubtful', *Philosophy of the Social Sciences*, **31**: 488–506.
Mantzavinos, C. (2001), *Individuals, Institutions, and Markets*, Cambridge: Cambridge University Press.
Rawls, J. (1995), 'Two Concepts of Rules', *Philosophical Review*, **64**: 3–32.
Smith, J. M. (1982), *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.
North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
Osborne, T. (1997), 'The Limits of Ontology', *History of the Human Sciences*, **10**: 97–102.
Ostrom, E. (1990), *Governing the Commons*, Cambridge: Cambridge University Press.
Parsons, T. (1935), 'The Place of Ultimate Values in Sociological Theory', *International Journal of Ethics*, **45**: 282–316.
Sawyer, R. K. (2002), 'Nonreductive Individualism Part I: Supervenience and Wild Disjunction', *Philosophy of the Social Sciences*, **32**: 537–559.
Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.
Searle, J. R. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Searle, J. R. (1990), 'Collective Intentions and Actions', in P. Cohen, J. Morgan and M. E. Pollack (eds.), *Intentions in Communication*, Cambridge, MA: MIT Press.
Searle, J. R. (1995), *The Construction of Social Reality*, New York, NY: 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: Oxford University Press.
Skyrms, B. (1996), *Evolution of the Social Contract*, Cambridge: Cambridge University Press.
Skyrms, B. (2004), *The Stag Hunt and the Evolution of Social Structure*, Cambridge: Cambridge University Press.
Sterelny, K. (2003), *Thought in a Hostile World: the Evolution of Human Cognition*, Oxford: Blackwell.
Sugden, R. (1986), *The Economics of Rights, Co-operation and Welfare*, 2nd edn, 2004, Oxford: Blackwell.
Tuomela, R. (1995), *The Importance of Us*, Stanford, CA: Stanford University Press.
Tuomela, R. (2002), *The Philosophy of Social Practices*, Cambridge: Cambridge University 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**: 337–369.
Vanderschraaf, P. (2001), *Learning and Coordination*, London: Routledge.
Weber, M. (1910), 'Diskussionsrede zu dem Vortrag von A. Ploetz über Die Begriffe Rasse und Gesellschaft', in *Gesammelte Aufsätze zur Soziologie und Sozialpolitik*, Tübingen, Mohr, pp. 456–462.
Wilson, E. O. (1998), *Consilience: The Unity of Knowledge*, London: Little, Brown and Company.
Ylikoski, P. and J. Kuorikoski (2012), 'Dissecting Explanatory Power', *Philosophical Studies*, **148**: 201–219.
Young, P. H. (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ: Princeton University Press.