

Understanding institutions: replies to Aoki, Binmore, Hodgson, Searle, Smith, and Sugden

FRANK HINDRIKS

Faculty of Philosophy, University of Groningen, Groningen, the Netherlands

FRANCESCO GUALA*

Department of Economics, Management and Quantitative Studies, Università degli Studi di Milano, Italy

Abstract. Our goal is to develop a theory that combines the best insights of philosophical and scientific theories of institutions. We are not committed *a priori* to save the commonsense notion of institution, or the thesis of human exceptionalism. We think that human cognition is important, but we do not claim that common knowledge or collective intentions are necessary for coordination. Like most of our commentators, we believe that there is continuity between simple rules of precedence and sophisticated institutions like property, marriage, or money. Finally, we argue that a satisfactory account of institutions must be compatible with different theories of normativity, specifying the social and psychological mechanisms that make it possible to override selfish desires.

We are grateful and honoured for the attention that six distinguished scholars have devoted to our paper (Hindriks and Guala, 2014). Their views have influenced significantly the position that we try to defend in the paper, and their commentaries will help us elaborate some points that require clarification. Before we do that, however, we would like to notice that an important thesis appears to be vindicated by these commentaries: there is a big gap between some philosophical accounts of institutions and those found in the scientific literature. This is particularly evident when we compare Searle's (2015) remarks with those of the other participants. While reading, we found ourselves frequently in agreement with the social scientists, and not by accident: our goal is to develop a unified theory of institutions that is informed by what the most advanced sciences (including the social sciences) have to say on this matter. Searle does not take the social sciences seriously, so unsurprisingly he finds our project misguided. We will address his qualms shortly.

According to Hodgson (2015), we ought to distinguish between two tasks: the task of defining and the task of understanding (or explaining) institutions. Hodgson believes that our unified theory does well on the explanation side, but

*Email: francesco.guala@unimi.it

is inadequate on the definition side. We agree only partly, although we find the distinction useful. One may seek definitions for many reasons: Hodgson points out that they facilitate communication, and argues that a good definition must isolate the essential properties of a kind. He also argues that rules are the best candidates for the role of essential properties of institutions. But Hodgson recognizes that not any rule will do. An ‘institutional’ rule must influence behaviour (‘act as a real constraint’), and one way to do it is by prescribing actions that are part of an equilibrium. Since equilibrium models are explanatory, it seems to us that there cannot be a neat divide between definitions and explanations. Science is in the business of explaining, and naturalists are committed to use the best science that is available, even when they give definitions.

Equilibrium models explain the continued existence of institutions – why they do not just exist, but also persist. This is important not only for understanding but also for policy-making, which is why social scientists have focused on equilibrium accounts of institutions. According to Binmore (2014), for example, equilibrium is the ‘big flea’ on top of which all the other fleas ride in the sense that, once you have equilibrium, you have most of what matters for understanding institutions. Binmore argues that a Nash equilibrium is more apt to appear in a theory of institutions than correlated equilibrium because in a Nash equilibrium the correlation device is explicitly stated, not ‘hidden from view’. We fully agree on the importance of correlation devices. The reason why we chose an exposition based on correlated equilibrium is that it highlights another important feature of institutions: the fact that coordination problems can usually be solved in different ways. There are many potential correlation devices, and although there are human universals – as Vernon Smith reminds us – there are also differences in the ways in which the same problem of coordination is addressed in different societies. Although we agree with Binmore that the unified theory could be expressed in terms of Nash equilibria, we chose this theoretical framework to highlight the variety of coordination devices.

This choice may have been misleading because correlated equilibria are often associated with Aumann’s other famous concept, common knowledge. Binmore thinks that common knowledge has been given too much prominence and that convergence on Nash equilibria should be explained using evolutionary models, rather than rational game theory. He is concerned that the rational approach (à la Aumann and Lewis) may lead to overlook the role of history, and may impose unreasonable epistemic conditions for successful coordination, like common knowledge of the signal sent by the coordination device (see e.g. Binmore, 2010).

As far as we are concerned, we share these worries only in part. History does play an important role in Lewis’ theory, although it is not exactly the same role as in evolutionary game theory. But Binmore is right that common knowledge cannot be necessary for successful coordination. Binmore thinks that the best way to avoid the common knowledge trap is to bypass cognition entirely, and to

focus on ultimate mechanisms like Darwinian selection. But one can recognize that people are not fully rational and at the same time believe that it is useful to understand the proximate causes that govern the evolution of institutions. Different explanations are legitimate at different levels of analysis, and we can zoom in and out depending on our explanatory goals.

On this matter, our position is close to Sugden's. Sugden (2015) does not think that psychology is irrelevant, but reminds us that it need not be complicated. Sophisticated reasoning – of the kind that requires collective knowledge or collective acceptance, for instance – is unnecessary for institutions to emerge and persist. Simpler cognition will usually do. Sugden makes these claims to criticize our approach, but the critique is based on a misunderstanding: we do not think that common knowledge is needed and we tried to be as non-committal as possible regarding collective intentionality, to avoid to get muddled in this controversial topic.

Sugden also reproaches us for requiring that the unified theory be consistent with commonsense ontology. He is not attacking a straw man, to be sure, because many philosophers do impose this requirement. On our part, however, we just believe that the unification of scientific and commonsense ontology is an interesting possibility to explore. Science can (and often does) depart from commonsense. In the case of institutions, we took consistency with commonsense as a working hypothesis rather than as a requirement of adequacy. The hypothesis, we think, turns out to be fruitful: when it comes to institutions, contemporary science and common sense are indeed by and large consistent with one another. But scientific accounts would not have been invalidated if the hypothesis had turned out to be false.

Finally, Sugden claims that a theory of institutions that applies across animal species is superior to one that does not: it is more general, and explains more with less. We agree, but contrary to what he suggests we are not committed to 'human exceptionalism' *a priori*. Perhaps the fact that ordinarily we do not speak of animal institutions is an accident that science will correct. But it is also possible that ordinary language reflects a significant difference between the causal mechanisms that govern human and animal societies. If so, it might make sense to talk of social practices among animals, but not of social institutions. This is another working hypothesis that we would like to put to the test.

How could the very human capacity to represent rules – highlighted by Aoki (2015), Hodgson (2015), and Searle (2010, 2015) among others – be significant? Well, suppose you want to intervene to change a convention. Depending on whether it is a human or an animal convention, you can do it in different ways, by pulling different levers. This is important both for science and for policy-making. Of course new animal equilibria can emerge in evolutionary time, as Sugden points out, but then you should better be aware that it will take a lot longer than the changes that occur in human societies. In the case of humans, it is possible to intervene effectively in a much shorter time span by manipulating

beliefs (representations). It is important to know this, but it is not something that you can know *a priori*, by analysing the commonsense notion of institution. It is an empirical fact, to be discovered *a posteriori*.

Binmore is right that there are larger and smaller fleas, and some differences between philosophers' and scientists' accounts of institutions depend on the way in which they assess the sizes of different fleas. We are grateful to Masahiko Aoki, Vernon Smith, and Searle for making this point emerge forcefully. Take normativity, for example: according to social scientists, it is a small flea. This does not mean that it is unimportant, or that it does not exist. But most social scientists think that it is not the fundamental property that Searle thinks. Smith (2015) mentions a nice example of socially unacceptable behaviour: queue jumping. Our first reaction, when someone jumps a queue, is normatively coloured ('it's unfair!'); but for a social scientist the first-come-first-serve rule is first and foremost an equilibrium, and an efficient one as well. This explains why the institution exists, persists, and is ubiquitous – a human universal as Smith points out. Our normative judgment from a scientist's perspective is just the icing on the cake: it can be explained from repeated interaction using insights from psychology and sociology. Rather than using normativity to explain the institution, social scientists like Aoki and Smith argue that the normative power emerges from the repeated play of equilibrium strategies.

So why does Searle see normativity as a big flea? Searle is committed to explain human uniqueness, and for his theory of institutions this is a requirement, not a working hypothesis. He thinks that human uniqueness has to do with the capacity to impose 'status functions', and that the latter involve 'deontic powers'. Notice that Searle's explanandum is extremely demanding; no one else thinks that an adequate theory of institutions must account for the eight features that he lists in his commentary. As we will explain shortly, we suspect that this list is a by-product of his philosophical commitments rather than of a genuine curiosity about institutions.

Searle is right to say that in our paper, we do not discuss normativity. We do not do it because this paper was originally aimed at an audience of social scientists; but we discuss normativity in a companion article published in the *Philosophical Quarterly* (Guala and Hindriks, 2014).¹ While inviting interested readers to have a look at it, here we will try to explain why we disagree with Searle on normativity and related issues.

Searle's refutation of the social science approach to institutions, and a fortiori of our unified theory, is based on three claims: (1) normativity is essential for institutions; (2) normativity requires constitutive rules; and (3) constitutive rules cannot be reduced to regulative rules. Bracketing the first claim, it is easy to show that both the second and the third claims are false.

1 Searle seems to have read it, since the example of the Nuer and the Dinka that he mentions appears only in that paper. Unfortunately, he does not engage with what we say about normativity there.

Normativity does not require constitutive rules: when Bob says to Ann ‘no one is allowed to use my garden without my permission’, he is uttering a regulative rule with normative force. He is not creating anything new, in the ‘X counts as Y’ fashion. Our thesis is that the kind of deontic powers Searle thinks are essential for institutions can be derived from these simpler normative rules. Searle disagrees, but it is not entirely clear why. Perhaps he thinks that regulative rules do not create rights and obligations, as they represent imperatives instead. It remains unclear to us, however, how the introduction of an institutional term like ‘property’ could have a creative power that imperatives lack. Imperatives *are* meant to express or even create obligations. When we say that Bob has the right to use his garden because it is his property, we are saying that Ann (and Charles, Diane, etc.) must not organize barbecue parties in it without his permission, that they cannot sell it without his permission, or rent it out without his permission . . . and so on. This reveals that the content of the institutional term ‘property’ can be translated into a set of regulative rules specifying a series of things that other people ought (not) to do. This translation is possible because my rights are your obligations and can be expressed by regulative rules.

So every attribution of rights is equivalent to stating a series of regulative rules. The difference between using these rules and using the term ‘property’ is just that the rules have been bundled in such a way that we can use a single institutional term to name them (the result is what we call ‘a status rule’, which we incorporate in the theory of constitutive rules to better bring out the connotation of institutional terms; Searle’s ‘X counts as Y’ formula leaves this implicit). Searle thinks that there is a huge gap between formal property rights and the impropriety of lifting people off their bus seats. But legal rights do not exist on a separate ontological dimension, as Adam Smith, David Hume, and Edward Burke explained long ago (‘the same mistakes apparently are being repeated today’). Bob Sugden and Vernon Smith in their commentaries explain why there is continuity between the assignment of bus seats and legal property rights, so we will not dwell on this point here.²

Another way to get at basically the same issue is by considering the examples that Searle uses to illustrate the notion of a regulative rule, the rules that require driving on the right in the US and driving on the left in the UK. Given that on Searle’s view institutions are to be analysed in terms of constitutive rules, he is committed to denying that these traffic rules are institutions. We see no valid reason for doing so. Interestingly, Searle points out that these rules of the road are legal requirements, but then goes on to deny that they generate deontologies in the way that institutions do. This can only mean that his conception of the normativity of institutions is too narrow.

² It is worth mentioning that this is not a peculiar view entertained by social scientists only; legal scholars also consider it a platitude (see e.g. MacCormick, 1998).

If both formal and informal institutions are normative, then, where does the normativity come from? Many social scientists, including Aoki, Binmore, Smith, and Sugden, have argued that it can emerge from repeated play. We find this account not at all implausible: in many cases it does seem to explain quite well what is going on.³ But we do not believe that this is necessarily the only story. Suppose we told you that you cannot park your car here. You might take us to imply that if you park your car here you will be fined; that if you park it here, you will block the way and annoy passers-by; that we will be annoyed; that we will form a bad opinion of you; that the old lady living there will not be able to open her door, and when you realise it you will feel guilty (and so forth).

Someone might object that these classic social scientific accounts explain only why people believe institutions have a normative dimension, but do not account for what that normative dimension really is. This distinction is controversial, but even if it holds, it is far from obvious that providing an account of ‘true normativity’ should be a core concern of a theory of institutions (the issue is discussed in Hindriks 2013). In developing our unified theory, we are in the first instance concerned with how the normativity of institutions affects behaviour. For this purpose, it suffices if we can explain why people believe institutions to be normative. And social scientific accounts of normativity try to do this by invoking factors such as punishment, internalized norms, and/or moral emotions.

Searle however finds these accounts unsatisfactory: in each case a desire (not to be punished, not to feel guilty) is lurking in the background, while genuine obligations, he thinks, are ‘desire independent reasons’. So, what should we make of this claim?

The notion of ‘desire independent reason’ can be given a weak or a strong interpretation. On the weak interpretation, obligations provide reasons to act that are independent of *some* desires (selfish ones, in particular). This interpretation is widely accepted – it is in fact a platitude – and it is entirely compatible with scientific accounts of institutions based on incentives and equilibria. It is also compatible with many different theories of normativity, which specify the social and psychological mechanisms that make it possible to override a selfish desire by uttering and accepting a normative statement.

On the strong interpretation, in contrast, obligations provide reasons to act that are independent of *any* desire. This is a much more controversial view: according to an influential tradition that goes back to Hume, all human actions – including actions that are performed to comply with norms – are driven by desires. The controversy between Humeans and their opponents is one of the central unresolved issues in the philosophy of action that continues to raise deep and important philosophical questions. Searle, however, has done little or nothing to answer them. His use of the term ‘desire-independent reason’ is

³ Guala and Mittone (2010) have investigated this mechanism empirically, and Hodgson *et al.* (2012) offer a speculative account – based on fMRI data – of the neural basis of social norms.

more a matter of putting a name to a problem, rather than of solving it. Thus, Searle faces a dilemma: either his critique fails to hit its target, or he is the one who has put forward a controversial claim that is in need of an argument and a concomitant theory. Meanwhile, we choose to focus on how the normative dimension of institutions affects behaviour, and believe that, instead of trying to adjudicate the issue, it is sensible to work with a concept of normativity that is neutral on the ‘true nature of normativity’ (if there is one).

Some philosophers believe that normativity can be explained in terms of mutual beliefs and the feeling of resentment that we experience when our expectations are frustrated (Bicchieri, 2006; Lewis, 1969; Sugden, 1998); others believe that normativity requires a stronger notion of collective 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 (Broome, 2013; Raz, 1999; Skorupsky, 2010). Whether any of these accounts is able to explain normativity in a satisfactory way is an open question that we do not want to settle here. Precisely because normativity is such a thorny issue, it would be foolish to make a theory of institutions depend on a specific account.⁴

Our aim has been to provide a unified theory that enables members of different traditions to see what they can learn from one another. It is clear that we have failed insofar as Searle is concerned. However, Searle admits candidly that he does not know the game-theoretic literature well; from what he writes it is clear that he does not know what a Nash equilibrium is;⁵ in a ranting paragraph on the prisoner’s dilemma, he even seems to ignore the basic difference between a scientific model and its interpretation. But his knowledge of the social ontology literature is also limited: contrary to what he says, the philosophers that he cites approvingly – like Margaret Gilbert, Geoffrey Hodgson, and Raimo Tuomela – are aware that their questions are intimately related to the ones that game theorists ask, and in their writings they have tried to build a bridge between the two literatures (e.g. Gilbert, 1989; Hakli *et al.*, 2010; Hodgson, 2006; Tuomela, 1985). This symposium confirms our belief that a bridge is badly needed, but it also reminds us that it will take a lot of effort to accomplish the task.

References

- Aoki, M. (2015), ‘Why is the Equilibrium Notion Essential for a Unified Institutional Theory? A Friendly Remark on the Article by Hindriks and Guala’, *Journal of Institutional Economics*, published online. doi: [10.1017/S1744137415000090](https://doi.org/10.1017/S1744137415000090).
- Bicchieri, C. (2006), *The Grammar of Society*, Cambridge: Cambridge University Press.

⁴ Incidentally, more than one account may be right because normativity may have different sources: there is no reason to believe that there is only one ‘true’ kind of normativity.

⁵ He seems to think that unequal allocations of wealth cannot be equilibria, for example.

- Binmore, K. (2010), 'Do Conventions Need to Be Common Knowledge?', *Topoi*, 27: 17–27.
- Binmore, K. (2014), 'Institutions, Rules and Equilibria: A Commentary', *Journal of Institutional Economics*, published online. doi: [10.1017/S1744137414000599](https://doi.org/10.1017/S1744137414000599).
- Broome, J. (2013), *Rationality Through Reasoning*, Oxford: Blackwell.
- Frank, R. (1987), *Passions Within Reason*, New York: Norton.
- Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.
- Gilbert, M. (1989), *On Social Facts*, Princeton: Princeton University Press.
- Guala, F. and F. Hindriks (2014), 'A Unified Social Ontology', *Philosophical Quarterly*, doi: [10.1093/pq/pqu072](https://doi.org/10.1093/pq/pqu072).
- Guala, F. and L. Mittone (2010), 'How History and Conventions Create Norms: An Experimental Study', *Journal of Economic Psychology*, 31: 749–756.
- Hakli, R., K. Miller, and R. Tuomela (2010), 'Two Kinds of We-Reasoning', *Economics and Philosophy*, 26: 291–320.
- Hindriks, F. (2013), 'Collective Acceptance and the Is-Ought Argument', *Ethical Theory and Moral Practice*, 16: 465–480.
- Hindriks, F. and F. Guala (2014), 'Institutions, Rules, and Equilibria: A Unified Theory', *Journal of Institutional Economics*, published online, doi: [10.1017/S1744137414000496](https://doi.org/10.1017/S1744137414000496).
- Hodgson, G. M. (2006), 'What Are Institutions?', *Journal of Economic Issues*, 15: 1–23.
- Hodgson, G. M. (2015), 'Much of the "Economics of Property Rights" Devalues Property and Legal Rights', *Journal of Institutional Economics*, published online. doi: [10.1017/S1744137415000028](https://doi.org/10.1017/S1744137415000028).
- Hodgson, T., F. Guala, T. Miller, and I. Summers (2012), 'Limbic and Prefrontal Activity during Conformity and Violation of Norms in a Coordination Game', *Journal of Neuroscience, Psychology, and Economics*, 5, 1–17.
- Lewis, D. K. (1969), *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.
- MacCormick, N. (1998), 'Norms, Institutions, and Institutional Facts', *Law and Philosophy*, 17: 301–345.
- Nichols, S. (2004), *Norms with Feelings*, Oxford: Oxford University Press.
- Raz, J. (1999), 'Explaining Normativity: On Rationality and the Justification of Reason', *Ratio*, 12: 354–379.
- Searle, J. R. (2010), *Making the Social World*, Oxford: Oxford University Press.
- Searle, J. R. (2015), 'On Status Functions and Institutional Facts: Reply to Hindriks and Guala', *Journal of Institutional Economics*, published online. doi: [10.1017/S1744137414000629](https://doi.org/10.1017/S1744137414000629).
- Skorupski, J. (2010), *The Domain of Reason*, Oxford: Oxford University Press.
- Smith, V. L. (2015), 'Conduct, Rules and the Origins of Institutions', *Journal of Institutional Economics*, published online. doi: [10.1017/S1744137414000605](https://doi.org/10.1017/S1744137414000605).
- Sugden, R. (1998), 'Normative Expectations: The Simultaneous Evolution of Institutions and Norms', in A. Ben-Ner and L. Putterman (eds.), *Economics, Values, and Organization*, Cambridge: Cambridge University Press, pp. 73–100.
- Sugden, R. (2015), 'On "common sense ontology": A Comment on the Paper by Frank Hindriks and Francesco Guala', *Journal of Institutional Economics*, published online. doi: [10.1017/S174413741500003X](https://doi.org/10.1017/S174413741500003X).
- Tuomela, R. (1985), *A Theory of Social Action*, Dordrecht: Reidel.