

REVIEW OF F. GUALA,  
*UNDERSTANDING INSTITUTIONS.*  
*THE SCIENCE AND PHILOSOPHY OF LIVING TOGETHER,*  
PRINCETON, PRINCETON UNIVERSITY PRESS, 2016

EMMA TIEFFENBACH\*

If one wishes to understand what money is, to whom should one turn as the most reliable source of knowledge? Of course, economists propose themselves as the experts on the matter. Who, if not those who study interest rates, prices and exchanges could know more about the nature of money? Yet, with a few exceptions, those philosophers in the burgeoning field of social ontology who ask ‘what is money?’ (or, for that matter, ‘what is a marriage?’, ‘what is ownership?’, ‘what is a cocktail party?’, etc.) tend to ignore what past and present social scientists maintain on these issues. John Searle comes to mind as having such an insulated approach. Since *The Construction of Social Reality*, Searle has repeatedly claimed to have found no adequate definition of institutions on the part of those scholars who are the most concerned with them.

*Understanding Institutions* leaves a different impression about the state of the art. In this elegant book, Francesco Guala shows that social scientists, economists in particular, can actually be trusted in how they understand the notions which they apply. Guala finds in game theory the relevant conceptual framework in which to explore the nature of institutions. Following David Hume, David Lewis, and Edna Ullmann-Margalit, he defends a view, termed the *rule-in-equilibrium* account, which conceives institutions as solutions to coordination problems. For example, to determine what a marriage is, we must look at the range of coordination dilemmas which married couples have resolved: Who cooks? Who picks up the children from school? Who washes the dishes? etc. Likewise, the institution of ownership states *who* will make use of *what*? The highway code indicates which side of the road to drive on. Cocktail parties? I gather they specify *when* to entertain and *who* to bond with.

---

\* Université de Genève. Address for correspondence: emma-tieffenbach@unige.ch.

The overall purpose of Guala's book is to show that a game-theoretic approach to institutions need not be at odds with more profuse philosophical accounts. Guala propounds a 'unifying account' that does justice to both sides. For all of its originality and insights, however, Guala's theory of institutions may be more reductive than truly integrative. As we shall see, those features of institutions that philosophers often stress – deontic powers, constitutive rules, the reflexivity of social kinds – do not fit well with his naturalistic approach.

The book consists of two parts. In the first, Guala presents and motivates his account, which combines two apparently unrelated views of institutions. On the one hand, the *institutions-as-rules* view conceives institutions as implemented rules that facilitate social relationships. Why do we follow these rules? The answer to this question is provided by the *institutions-as-equilibria* view, which stresses the interest that people have in finding a solution to coordination dilemmas. The problem with the *institutions-as-equilibria* view is that it is too inclusive. The situation in which, in times of drought, residents keep hosing their lawns is a paradigmatic case of a Nash equilibrium. And yet it does not correspond to any institution. Equilibria can also be found in animals, but the latter do not have *bona fide* institutions. Guala's *rule-in-equilibrium* account retains the strength of the two accounts. At its core is the notion of 'correlated equilibria', which are equilibria upon which agents would converge if they were helped by a 'public event' (a ceremony, a decree, or any environmental cues making certain options focal) that serves as a coordination device. Public events index certain strategies on certain conditions: "if in Commonwealth countries, drive on the left, if elsewhere, drive on the right". Because the *rule-in-equilibrium* view conceives institutions as ruled-governed actions, rather than as strategic choices, it captures the normative aspects of institutions that are missing from the *institutions-as-equilibria* view.

In chapter five, Searle's theory of institutional facts in terms of constitutive rules – of the 'X counts as Y' form – is presented as one further instance of the *institutions-as-rules* view. This classification is surprising because the rules that pertain to the *institutions-as-rules* view are regulative ones. However, drawing on earlier works with Frank Hindriks, Guala contends that constitutive rules can be eliminated. The argument turns on an explication of the Y terms as referring to rule-governed activities. For example, private property is whatever certain person can make use of. Money is whatever entity is acquired for the sake of exchanging it. Constitutive rules are thus reduced to regulative rules by treating the Y institutional terms as shortcuts for more lengthy descriptions of particular activities.

In the 'interlude' that separates the two parts of the book, Guala discusses, first, the topic of mindreading. Following David Lewis, he conceives

coordination in terms of mutual expectations, while also rejecting Lewis' notion of common belief as cognitively too demanding. As Guala suggests, lower-level cognitive capacities, like empathy, imitation, or replication of feeling, are all that is needed for a meeting of minds. The interlude also discusses the need for collective intentionality as a necessary feature of institutions. Contrary to Gilbert, who sees in joint commitments an essential feature of social conventions, Guala notes that racist institutions involve no joint commitment on the part of those against whom they discriminate. One may reply, however, that while Gilbert's account of conventions is ill-suited to explaining racist institutions, Guala's rule-in-equilibria approach may not fare better in this respect. Victims of discriminatory institutions display behaviours (where they sit in a bus, where they reside in cities, who they marry) that others have imposed on them. The notion of equilibrium does not account for the coercive aspect of discriminatory institutions, and in fact misconstrues them as solutions to strategic games.

The second part of the book draws the implications of the *rule-in-equilibrium view* in regard to several philosophical debates in social ontology. These chapters represent, in my opinion, the most innovative, captivating and thought-provoking part of the book. In the chapter "Reflexivity", Guala revisits the notion of self-fulfilling prophecies as further instances of rules-in-equilibria. In light of paradigm examples (children performing as badly as they are expected to; a run on a bank), he first recognizes that classifications do modify institutional kinds. As he puts it, "the observation of behaviour confirms the expectations of the players, and the expectations induce behaviour that is consistent with the expectations" (125). But Guala also shows that once self-fulfilling prophecies are construed in this way, there is nothing intrinsically bad (nor, for that matter, intrinsically good) about them. For instance, the situation in which bank clients are when they all keep their money stored in the bank is just another, albeit desirable, instance of a self-fulfilling prophecy: the choice is one that nobody has an incentive to change unilaterally.

Chapter 10 discusses the distinction between natural kinds and social kinds, and the tendency, examined by social constructionists, to offer pseudo-explanations that deny such a difference. For instance, it is easy to believe that the 'woman' category is a natural kind. But many of the traits that we believe to be intrinsic to women (such as being good at raising children) are not biologically based. If these traits occur more often in women, it is in response to our sexist expectations. According to Guala, the rules-in-equilibrium view of institutions casts a new light on the underlying causal mechanism at work. Indeed, the theory reveals that certain social arrangements (he works/she raises the children) are "not inevitable, and the reason is that they are just one equilibria among others". So social

constructionism and the rules-in-equilibrium view similarly stress the contingency of institutional rules.

Chapter 11 deals with the view that institutions non-causally depend on our representations for existing. The idea that money is what is believed to be such is a leitmotiv in social ontology. Versions of this idea can be found in Ruben, Thomasson, Dilthey, Gadamer, Ricoeur, Taylor, Peter Winch and, of course, Searle. Guala stresses three main implications of this view, namely, that social kinds are not real, that we cannot be possibly wrong about them, and that they can be known from the armchair. These three implications are then refuted in chapter twelve, where Guala convincingly argues that what makes some entity  $x$  (a piece of paper) a member of a social kind  $s$  (money) is not fixed by people's representation of the conditions (say, being issued by the central bank) that make  $x$  a member of  $s$ . The real content of social terms, Guala shows, is determined by the *actions* or patterns of activities associated by these terms. A party is the action of gathering and having fun. Money is the action of accepting the same thing as means of payment. The upshot of the analysis is that a social institution is "a system of actions and expectations" (172) that exist independently from folk-psychological views about its condition of existence.

Guala's definition of institutions ultimately rests on the identification of their typical functions, but how are these to be discovered? Guala is no friend of armchair *a priori* explorations. The properties of social kinds, he claims, are "discovered *a posteriori*, through empirical investigation". But, as he also admits, relying on empirical studies begs the question as to how genuine and pseudo instances of social kinds must be distinguished. However uncommon same sex unions have been, it seems wrong to rule them out as pseudo-examples of marriage. To reconcile realism (institutions are worldly determined) with reformism (institutions can be substantively reshaped for the better), Guala urges a functionalist view: whichever form they take, the conceptual boundaries of institutions are to be fixed by the coordination problems which they solve.

The above summary blatantly fails to cover the range of issues and examples that Guala manages to consistently integrate in the 14 chapters that compose his book. *Understanding Institutions* painlessly introduces novices to the most important issues in social ontology. Guala is obviously good at making sense of a wide range of situations by reading them through the lens of game theory. I am thinking, among many others, of the passage where he applies Thomas Schelling's S-shaped model of propagation to explain why an increasing proportion of people in a population choose to 'come out'. I suspect that the same mechanism may be at play in the more recent #metoo movement. While admiring the achievement of making a wide range of issues in social ontology accessible, the experts will also find

much to reflect upon—and potentially challenge. In the remaining part of this review, I will point to three concerns.

First, Guala construed institutions as having coordination dilemmas as pre-conditions. Interestingly, Carl Menger's explanation of money, which Guala invokes in support of this view, may in fact not perfectly illustrate it. Granted, Menger's account shows that a central authority need not play a coordinating role in helping agents converge on the same good. But the story may also be interpreted as showing that *agents need not coordinate at all* for money to arise. The few barterers who start using some goods as media of exchange only need to ask themselves, when reviewing candidates for that role, "how popular is this good on the market place?". And while the barterers need to anticipate correctly the saleability of, say, cowry shells, before ascribing them the function of media of exchanges, the point is that they need not be concerned by the reason(s) why cowry shells sell well. This is because whenever someone uses cowry shells as a means of exchange, s/he 'votes' for them, so to speak, thus further increasing their saleability on the market place, in a self-reinforcing manner. If, in the end, cowry shells perform by and large the function of money, this is due to a snowball effect whereby the uncoordinated choices of a few barterers influenced the choices of others, and so on until all unintentionally converged on cowry shells. The crux of Menger's story is, as I read it, to dispense with all coordination devices, be they centralized or decentralized. But if this is so, it follows that it is too strong a requirement to ask of institutions that they solve coordination dilemma. For a coordination dilemma may not characterize all pre-institutional stages.

My second critical remark concerns Guala's treatment of institutional normativity. According to Guala, the equilibrium approach that he favours supports a functionalist theory of institutional norms that explains their existence by pointing to their benefits. Norms exist because they make certain choices more predictable, if only by making certain options more salient. The idea is that because people have an interest in solving coordination dilemmas, they have an incentive to follow the norms that help them converge on the appropriate equilibria. But many have objected to this cost and benefits analysis of institutional rules, suspecting that it does not account for all the complexities and varieties of institutional 'oughtness'. There is, first, the problem, put forth by Jon Elster, of pointing to something that occurs at  $t_2$  in order to explain something that occurs at  $t_1$ . That the effect of institutional rules is to enhance coordination cannot retroactively explain why these rules were adopted in the first place, unless one provides the feedback loop that links the effect to its cause. Another shortcoming of reducing rules to incentives is that it fails to capture the nature of their binding force. Guala considers making others' behaviours

more predictable as one prominent way in which institutional rules enhance coordination. But it is not certain that institutional rules weigh on us as mere predictive tools. As H. L. A Hart notes in *The Concept of Law*, red lights indicate that drivers have an *obligation* to stop, and not only that they *are likely* to do so. The non-instrumental notion of obligation at play here does not fit well into Guala's functionalist framework.

My third critical comment regards the convergence that Guala finds between social constructionism, on the one hand, and his own account, on the other. Both views, Guala contends, show that social arrangements are 'not inevitable', but the argument, I submit, rests on an ambiguity regarding what it means for a social arrangement to be 'not inevitable'. In the context of equilibrium theory, the term means that the social arrangement is just one solution, among equally rewarding ones, to a coordination dilemma: the left side of the road is as good as the right side for driving. Now social constructionists seem to refer to social arrangements as 'not inevitable' in a different sense. What they mean is that certain social arrangements are the products of cultural factors that are partly under our control and that therefore could be otherwise if only we freed ourselves from our biased expectations. Whereas the game theorist stresses the equivalence (in terms of pay offs) of options, the social constructionist focuses on our responsibility in letting these options prevail as the result of our prejudicial expectations about *who* must do *what*. Put differently, the social constructionist is likely to dispute the game theorist's assumption that the two options 1) 'she raises the children/he works' and 2) 'he raises the children/she works' are equivalent. Unlike the game theorist, s/he will also insist on the importance of spelling out the details of the contingent path that widely resulted in 2).