Three conceptions of a theory of institutions

N. Emrah Aydinonat[1] & Petri Ylikoski[2]

Abstract: We compare Guala's unified theory of institutions with that of Searle and Greif. We show that unification can be many things and may be associated with diverse explanatory goals. We also highlight some of the important shortcomings of Guala's account: it does not capture all social institutions, its ability to bridge social ontology and game theory is based on a problematic interpretation of the type-token distinction, and its ability to make social ontology useful for social sciences is hindered by Guala's interpretation of social institution types as social kinds akin to natural kinds.

Keywords: Institutions, unification, social ontology, economics

Introduction

Francesco Guala develops a unified theory of institutions in his book *Understanding Institutions*. He presents his proposed theory, which unites the rules and equilibrium approaches to institutions, as an alternative to John Searle's account. He synthesizes the insights from various fields to establish the continuity between social ontology and social sciences, and to gain a better understanding of institutions. In the present paper, we compare Guala's project with Searle's theory and Avner Greif's attempt to create a unified framework for studying institutions. We chose the first for the obvious reason that Guala's and Searle's projects are rivals within social ontology. Greif's framework, on the other hand, was chosen for two

[1] Corresponding author: N. Emrah Aydinonat, Social and Moral Philosophy, Faculty of Social Sciences, University of Helsinki, PO Box 24, 00014 Helsinki, Finland. E-mail: aydinonat@gmail.com
[2] Sociology, Faculty of Social Sciences, University of Helsinki.

reasons. First, it is noticeably similar to Guala's project, not only in that it combines the rules and equilibrium approaches, but it also brings together insights from various fields. Second, the subtle differences between the two help us to consider the extent to which Guala's theory could be useful from a social scientific perspective.

A comparison of these three views of institutions not only illuminates Guala's project, but also serves to illustrate that a unified theory of institutions can be many things. Our primary interest is in how these three theorists conceptualize unification, the appropriate level of abstraction, and the explanatory goals of the theory. In other words, we will compare their projects advocating a unified theory of institutions so as to learn some general lessons, not to assess how successful they are in accomplishing their goals. The latter task would require a much longer paper. Throughout the paper we illustrate the differences and similarities between these three accounts of institutions with respect to unification, level of abstraction and explanatory goals. Based on this comparative analysis we also point out some of the shortcomings in Guala's account of social institutions. First, we argue that the proof-of-the-concept nature of Guala's theorizing prevents him from producing an all-encompassing account of social institutions. Complex social institutions such as 'institutions', 'university', 'government' does not fit comfortably to Guala's rules-in-equilibrium approach. Second, we point out that Guala's type-token distinction plays a central role in his attempt to unify social ontology and to bridge game theory and social ontology. We show that this distinction is problematic and for this reason it remains unclear whether Guala's account can capture the "essence" of any particular institution—in Guala's terms, of token institutions. Third, Guala defines institutions, such as marriage, on the bases of the coordination and cooperation problems they solve, and the social activities they regulate. However, he is also aware of the fact that any two particular examples of an institution might not solve the same set of problems or regulate the same set of social activities. To overcome this difficulty, Guala presumes that social institution types can be conceived of as social kinds that are analogical to natural kinds. He further assumes that the

general concepts that social scientists use—such as 'marriage'—overlaps with these natural kinds. We argue a that this is a dubious solution to the problem of general concepts in the social sciences. Fourth, we highlight that Guala's account promises to produce a general understanding of stability and persistence of institutions, but not their emergence. We argue that Guala's functionalist approach to institutions lacks the components that would be helpful for social scientists in reaching a better understanding of *particular* (token) institutions. Finally, we argue that the study of institutions should focus on particular institutions as historical causal complexes, not on their postulated abstract functional essences and suggest that social mechanisms should constitute the generalizable core of the social sciences.

The paper is organized as follows. The first section highlights some central features of Guala's theory. The second section compares it to Searle theory. The third section introduces Greif's unified framework and contrasts it with Guala's unification attempt. The fourth section reflects on what can be learned from the comparison and considers the implications for Guala's project. The final section concludes the paper.

Guala's unification project

Guala (2016) frames his approach in *Understanding Institutions* in terms of unification, beginning with the title of the first part of the book. Unification, in its different senses, is indeed an important characteristic of his approach. It is thus necessary to understand the senses in which his theory unifies. He appears throughout the book to be discussing three related conceptions of unification. First, he argues that "social ontology has been disunified for too long, and it is time that we put it together again" (2016, xxx), and he sets his main aim as "to unify the main traditions in the field of social ontology" (2016, xix). This is the book's first promise: the unification of social ontology. Second, Guala attempts to unify the accounts of institutions in the social sciences and social ontology: let us call his attempt to unify these two fields *field*

unification. This is the second promise of the book. Guala's intellectual strategy for delivering these promises is to bring together theoretical ideas that have been developed separately, and to synthesize them to build a general theory. Let us call this the *composite unification strategy*. We will now take a closer look at these ideas about unification.

The context for Guala's project of field unification was conveniently set by John Searle. As Guala informs the reader, John Searle "has stated boldly that he cannot find anything helpful in the *entire* social science literature". He goes on: "the dissatisfaction is reciprocal: social scientists have struggled to appreciate what the contribution of philosophy might be" (2016, xviii). Guala finds this situation unsatisfactory. He argues that although different scholars might "have different concerns", their questions are generally the same: "What is an institution? What is the social world made of? How many kinds of social entities are there?" (2016, xxx). Thus, he continues to argue, one must find out whether the different answers provided by philosophers and social scientists are compatible, and if so, to what extent. This sets the task for the book. The aim is to demonstrate the possibility of field unification by presenting a theory that unifies "the main traditions in the field of social ontology" (2016, xix). Hence, the unification of social ontology is the path to field unification. It is for this reason that we need to consider how Guala delivers this promise.

At this point the strategy of *composite unification* becomes relevant. Guala's argument could be interpreted as a *series of* composite unifications. First, he suggests that we can unify two broad rational-choice approaches to institutions (namely, rules and equilibrium approaches) in the more powerful *rules-in-equilibrium* approach (2016, xxiii). He argues that this approach unifies the two approaches found in economics, suggesting that institutions "are better conceived as rules that people have an incentive to follow" (2016, 10). Second, he suggests that a number of ideas developed in social ontology —Frank Hindriks' reductive theory of constitutive rules (Hindriks and Guala 2014; Guala and Hindriks 2015; Hindriks 2005, 2009), David Lewis'

(1969) account of conventions, and the incentivized action theory (Smit, Buekens, and du Plessis 2014, 2011) — could be combined with the *rules-in-equilibrium* approach. Hindriks' argument allows him to reduce Searle's constitutive rules to ordinary regulative rules. This makes it possible to connect social ontology to Lewis' account of conventions, and to game-theoretical accounts of institutions that it inspired—without the need for constitutive rules. Incentivized action theory makes it possible to link reasons for action with game-theoretical ideas, and Lewis does the same with conventions. These moves allow Guala to discuss institutions using the toolset of game theory, and to conceptualize them in terms of the rules and equilibria of games. His principle innovation does not concern these ingredients, but it is in how he puts them together. His main argument is that the resulting composite theory—or at least the outline of it—demonstrates the possibility of field unification, is stronger than its individual components separately, and, finally, has more explanatory power than earlier field-specific theories. This argumentative strategy explains the prominence of Searle in Guala's discussion: if he can successfully bring together ontological ideas that have been developed as alternatives to Searle's views (i.e. unifying the opposition) and combine them with social-scientific concepts, not only will he have produced a comprehensive alternative to Searle's social ontology, he will also have demonstrated that Searle's posit of the separation between the fields is wrong.

Guala's unification project clearly has an explanatory motivation. He explicitly expresses this motivation in a paper written jointly with Frank Hindriks, in which they claim that the "payoff of unification is a significant increase in explanatory power" (Hindriks and Guala 2014, p. 469). Guala argues against rules-based accounts, for example, observing that they do not explain why people follow the rules. Combining the rules-based account with the equilibrium account, makes it possible to explain why people are motivated to follow the rules that constitute the institution. Furthermore, he argues, it explains why the institution may become persistent. In this way, the unified theory provides a better explanation of the *raison d'être* and persistence of institutions.

Nevertheless, there are limits to what a general theory can do. Guala recognizes that the rules-in-the-equilibrium account is not sufficient to explain the emergence of an institution because it presupposes the existence of a correlating device. In the presence of multiple-equilibria, explaining emergence requires the studying of particular institutions and their history.

> "[...] equilibrium models provide legitimate explanations of the persistence of behavioral patterns. This does not mean of course that equilibrium models are well suited to provide other types of explanation, or that persistence is the only kind of phenomenon worthy of investigation in the social sciences. On the contrary, there are many other interesting aspects of social reality that call for different styles of explanation. But simple game-theoretic models can illuminate many features of institutions that otherwise would appear rather incomprehensible." (Guala 2016, 31)

Guala presents an appealing way of building a theory of social institutions. Before we compare it with alternative visions of a unified theory of institutions, we will make some observations about the theory his composite strategy has produced. This theory unifies social ontology by bringing together existing ideas from social ontology and the social sciences. Guala's grand claim is that the theory accomplishes field unification: it unifies social ontology and social scientific theorizing about institutions. However, what he has done is to connect social ontology with game theory-based theorizing about institutions. It remains to be demonstrated whether Guala's theory can be used to reconstruct social-scientific theories and ideas that do not use game theory. Thus, complete field unification remains a promissory note in the book. It is still to be shown that the bridgehead of economic modelling can be extended to the deep inlands of social-scientific theorizing.

There is another promissory note that concerns the general applicability of the theory. Guala merely discusses some highly-stylized paradigmatic examples of institutions, such as money,

property rights, and marriage. It remains to be seen whether or not the theory can be applied to detailed examples of these institutions, or to institutions that are somewhat different from these core cases. For example, many institutions, such as 'government' and 'university', are highly complex and the ability of Guala's account appears to be severely hindered by the rules-in-equilibrium approach. At the very least, it remains an open question whether such complex institutions can be fully reconstructed within the rules-in-equilibrium framework. We are not arguing that this programmatic nature of the theory is a problem in itself – all the alternative theories face the same challenge – but it is important to recognize the proof-of-the-concept nature of Guala's theorizing.

The lever that allows Guala to bridge social ontology and game theory is the idea about levels of abstraction. For him there are three levels of abstraction in the analysis of institutions. The first is the level of particular institutions. These *token institutions*, as he calls them, have a history and a geographical location (Guala 2016, xx). The social sciences consider institutions at more general level, being concerned with *types of institution*: "monarchy", "political party", "family", and "golf club" are some of his examples. Theories and models about types of institutions can differ in terms of generality: theories about Churches are more general than theories about Protestant Churches, for example. On the most general level the social sciences meet social ontology. According to Guala

> "Social ontologists investigate the social world at the highest level of abstraction, and devise theories that are supposed to hold for all social institutions irrespective of their individual features." (Guala 2016, xxi)

This is the trick: by conceptualizing the social sciences and philosophy in terms of levels of abstraction, Guala finds a meeting point between the two disciplines. Taking the common denominator of all institutions to be their function (institutions "facilitate coordination and cooperation"; 2016, xxiii), he conceptualizes all institutions as having a common essence:

7

> "What all the institutions share is that they are solutions to the same
>
> problems, or equilibria of the same class of games." (Guala 2016, 196)

The hierarchy of levels of abstraction plays a vital role in Guala's account. All particular

institutions are tokens of social-scientific institution types. Social ontology in turn, deals with

the highest level of abstraction: it is concerned with what is common to "all these types and

tokens" (Guala 2016, xxii). This is not as simple as it sounds: consider what he says about

marriage.

> "So by studying marriage practices in Florence during the thirteenth century,
>
> say, we can only learn about the particular way in which Florentine people
>
> organized child-rearing, reproduction, and economic cooperation at a
>
> particular time. To find out what marriage is *in general*—what marriage is *as a*
>
> *type*—we need to focus on the coordination problem as a whole, rather than on
>
> its specific solutions, because institution types are defined functionally, by
>
> reference to their goal or to the problems they solve. Since the same goal can
>
> be attained in many ways, the study of specific institutions tells us more about
>
> the specific solutions than about what they have in common. What all the
>
> institutions share is that they are solutions to the same problems, or equilibria
>
> of the same class of games." (Guala 2016, 196)

We find the loose way in which Guala employs the type-token distinction problematic. All

Florentine marriages during the thirteenth century are tokens of the type 'thirteenth century

Florentine marriage'. However, it is not as obvious that the Florentine institution of marriage is

a token of the more general type 'marriage'. The latter is an observer's category (or a theoretical

concept) that lumps together similar enough institutions from different societies (and time

periods). There is no guarantee that it captures the functional essence of all marriages, or even

that there is such an essence. What we are saying is that both Florentine marriages and the

institution of the Florentine marriage are *instances of the concept* marriage, but they are *not tokens of the type* 'marriage'. The reason this sophistry matters is that the former relation is much less demanding than the latter. All particular Florentine marriages presuppose the institution of Florentine marriage. They would not be Florentine marriages without there being the type, in other words the Florentine practices of arranging reproduction, inheritance and so on. However, the existence of the institution that we call Florentine marriage does not presuppose the existence of the more general type. Furthermore, the concept of marriage is a family-resemblance term, which implies that marriages need not have a shared essence. As an anthropological concept marriage is a flexible interpretive notion that does not allow general definition (Sperber 1996). Furthermore, the same point applies to many other general institutional concepts, including money and property. This has serious implications because Guala's account of theorizing rests on a strong assumption about the relation between social-scientific concepts and particular social institutions. If the type-token relation does not hold, there is much less reason to expect that abstract game-theoretical models can capture the functional essence of any particular institution. Furthermore, it implies that theories of social ontology cannot be tested "against types of institutions" (Guala 2016, xxi), and that they must be tested against facts about particular institutions. We will return to the implications of this observation in the final section.

In sum, Guala presents us with an alternative ontological reconstruction of institutions that also explains their stability and persistence, but not their emergence. The explanation of emergence, and of change more generally, requires a different explanatory approach. The acknowledgement of this limitation brings up the central question of this paper: are ontological reconstruction and the explanation of persistence the main goals that a unified theory of institutions should aim to achieve?

Guala and Searle

Guala's main adversary with regard to his book is John Searle. Quite clearly, Guala is attempting to provide an alternative to Searle's vision of social ontology. This is reflected in substantial disagreements on issues such as the necessity of constitutive rules, the nature of deontic powers, and the need for collective intentionality. However, there are also deeper differences that reflect the fact that the two authors are engaged in quite different philosophical projects. Guala does not merely present an alternative ontological reconstruction of social institutions, he also provides an alternative vision of the relation between social sciences and philosophy, and even more importantly, a different account of the tasks of philosophical theory.

Searle's project of social ontology is purely philosophical. The main motive is to fit social reality to a more general ontology (Searle 1995, 5). The aim is not to integrate his account of social institutions into what social scientists are saying, but to make it compatible with other areas of ontology. In other words, he is concerned with how social things "hang together" with all the non-social stuff in the world. Consequently, his main task is to address these features of the social world that seem anomalous from the perspective of non-social ontology. The unification he seeks, is achieved in a general philosophical theory. Accordingly, his strategy is to provide a general theory about the ontology of institutional reality, and to show that this theory can be accommodated with a more general philosophical theory.

This general philosophical goal has some important consequences. First, Searle's account of institutions is not primarily explanatory: it rather concerns the ontological reconstruction of the institutional reality. One of his most central philosophical claims is the following:

> "With the important exception of language itself, all of institutional reality [...]
> is created by speech acts that have the same logical form as Declarations."
> (Searle 2010, 12–13)

His concern with "the logical form" of social reality explains why Searle seems so unconcerned with questions such as how people manage to come to share the same collective attitudes, how institutional change occurs, or how institutions persist. Social-scientific concerns with explaining institutional emergence, change, and stability are not salient for Searle: ontological issues have to be settled first, and only then can social scientists start their work. In Searle's view, the relationship between philosophical and social-scientific theories is asymmetric: foundational philosophical theory sets a constraint on acceptable social-scientific theories, but social-scientific theories do not have a similar role in philosophy. This is quite different from Guala's position, which is that philosophy and the social sciences are continuous and thus positioned more symmetrically. In Guala's view, for example, philosophical theory of institutions should be able to account for their stability and persistence while leaving their emergence and change to social scientists. Searle ignores all three explanatory concerns.[3]

It should be noted that Searle's disdain of the social sciences does not follow directly from his philosophical goals. He could have chosen to pursue social ontology using social-scientific materials and examples. Although the social sciences are mostly continuous with everyday thinking, there are elements that expand it. They have much to say about macro scale social reality, for example, whereas Searle dismisses macro social facts as mere "systematic fallouts" (Searle 2010, 22). There is no deep philosophical justification for this kind of arrogance and lack of intellectual curiosity. Nevertheless, it is possible to conceive of Searle's philosophical project in more social-science-friendly terms. We should not dismiss it merely because it does not seem to connect with the social sciences. For this reason, Guala is fully justified in addressing Searle's main philosophical claims head on.

---

[3] For the kinds of questions that Searle thinks a theory of institutions must answer see Searle (2015, 508–9).

It is nevertheless possible to provide arguments that directly challenge the viability of Searle's philosophical project. Consider, for example, the role of language in Searle's and Guala's theories. For Searle, language is the most fundamental institution. We think he is right in this. Language has a crucial role in the creation of institutional facts. Furthermore, linguistic conventions, grammatical rules, meanings, and speech acts are prime examples of social institutions. However, Searle fails to make his theory general in that he does not provide a full account of language as an institution, for the most part seemingly taking it as a given. Even more surprisingly, his theory is somewhat circular: as the above quotation shows, declarative speech acts are the basic building blocks of his theory. More precisely, he claims that institutional reality and declaratives have the same logical structure. However, this seems more like an intermediate result than a fundamental starting point. After all, it makes sense to ask what kind of social institution a declarative speech act is (Smit and Buekens 2017). Only when the answer comes could we have building blocks for a more general account of social reality. Thus, although Searle includes language within the scope of his theory, he fails to deliver. Guala's case is in many ways the opposite in that the announced scope of his theory does not include language. Consequently, it seems less unifying as it is narrower in scope. This is a pity, because in our judgment his theory is better equipped to handle the institutional elements that are constitutive of language. Furthermore, explicit consideration of language would have helped Guala to formulate applications of his theory. For example, his discussion of 'meaning' (Guala 2016, Chapter 13)—which is somewhat disappointing—would have benefited from this.

Guala and Greif

We have shown that Guala's unified theory combines the rules and equilibrium accounts of institutions. Although economists sometimes characterize institutions as "rules of the game"

(North 1990; Ostrom 1990) and sometimes as equilibrium phenomena[4] (Calvert 1995; Schotter 1981; Sugden 1986), they commonly recognize that these views are complementary, and acknowledge their limitations.[5]

> "Which definition of an institution to adopt is not an issue of right or wrong; it depends on the purpose of the analysis" (Aoki 2001, 10).

> "These two approaches can be seen as complementary parts of the analysts [sic] toolkit" (Greif and Kingston 2011, 41).

There have been attempts to bring the rules and equilibrium accounts under an inclusive framework to study various aspects of institutions (e.g., Aoki 2001; Greif 2006). The most prominent of these is Avner Greif's attempt to create a unified framework by synthesizing various definitions of institution found in the social sciences literature. In *Institutions and the Path to the Modern Economy* (2006), Greif defines an institution as "*a system of rules, beliefs, norms, and organizations that together generate a regularity of (social) behavior*" (Greif 2006, 30). He calls the components of this system—i.e., rules, beliefs, norms, and organizations—*institutional elements*. According to his characterization each institutional

---

[4] Occasionally they also distinguish between the players of the game and the rules of the game (North 1990) to separate organizations such as firms, banks, and regulatory bodies from the formal and informal rules that enable and restrict their actions.

[5] It is a mistake to assume that any one of these approaches is uniform in terms of its methodology. For example, equilibrium view of institutions employs a diverse set of equilibrium concepts (Nash, subgame perfect, evolutionarily stable, stochastically stable, etc.) and assumptions (full rationality, bounded rationality, perfect information, imperfect information etc.). These different approaches are also considered as complementary by many economists (e.g., Aoki 2001, 194–97).

element "is social in being a man-made, nonphysical factor that is exogenous to each individual whose behavior it influences. Together these components motivate, enable, and guide individuals to follow one behavior among the many that are technologically feasible in social situations" (Greif 2006, 30).

Greif's motivation for unifying these two approaches is similar to Guala's, who also argues that the rules approach alone cannot explain why people follow certain rules, but not others. However, Greif's attempt to unify the definitions of institutions is not restricted to these two views. His definition "encompasses other seemingly alternative definitions" and aims at fostering the "integration of insights and analytical frameworks developed in conjunction with various definitions of institutions" (Greif 2006, 29). In fact, he purports to bring together different traditions in the social sciences and economics: old and new institutional economics, as well as prominent accounts of institutions in sociology, organization studies, history, political science, and philosophy (Greif 2006, 22–23, 39–40).[6] In contrast with Guala's account, this list is much more extensive. It is also noticeable that although social ontology is included, it has no special role. It is just one approach that might have some insights into the nature of institutions.

Greif's unification attempt might seem overly ambitious, but he does not seek to absorb all of these distinct approaches in one comprehensive theory. Moreover, from a purely philosophical perspective, his definition of institution may appear circular: a philosopher might regard many of his institutional elements as institutions in themselves, for example. However, Greif's aim is not ontological reconstruction, but rather to build a common analytical framework for explanatory purposes. Furthermore, his explanatory targets are *particular* institutions. He describes his project as an "attempt to gain a better understanding of a particular historical episode and to learn about institutions in general from this period" (Greif 2006, xvi). He argues

---

[6] Among others Greif cites Searle (1995).

that the unified framework should not accept generalizing presumptions—concerning the origins, functions and effects of institutions—of any of the distinct views his synthesis brings together. He even suggests that any *theory* of institutions is limited in its ability to study their origins, functions and effects because of their essential indeterminacy: institutions are complex, contingent and context-dependent phenomena, and understanding particular institutions requires theoretical as well as empirical/historical research (Greif 2006, 350–76).

Guala and Greif see the role and prospects of a general theory quite differently. Although both accept the context-specificity of institutions, unlike Guala, Greif considers this a constraint on what the theory can accomplish. Whereas Guala is optimistic about finding commonalities between institutions at a very general level, Greif places his bets on the usefulness of theoretical ideas in enhancing understanding of particular institutions. His approach is a refined example of the *analytic narratives* approach (Bates et al. 1998), which claims to take the context-specificity of institutions seriously.[7] As a consequence, Greif considers game theory a useful analytical tool (Greif 2006, 387), but refrains from defining institutions as game-theoretic equilibria. He argues,

> "Institutions are not game-theoretic equilibria, games are not the basic unit of
> institutional analysis, and game theory does not provide us with a theory of
> institutions. Indeed, the key to advancing institutional analysis by using game

---

[7] An analytic narrative combines "analytic tools that are commonly employed in economics and political science with the narrative form" that "pays close attention to stories, accounts, and context". Analytic narrative is, nevertheless, "analytic in that it extracts explicit and formal lines of reasoning" that draws predominantly on game theory and rational choice theory to "facilitate both exposition and explanation" (Bates et al. 1998, 10). Greif argues that this approach is particularly suited to the study of institutions because of the context-specificity of institutions.

theory is precisely to recognize the difference between game-theoretic equilibrium analysis and institutional analysis." (Greif 2006, 19)

Institutional analysis, unlike abstract game-theoretical analysis, must take the local and historical institutional constraints into account. Greif is not solely focused on explaining the stability or persistence of institutions, he is also interested in accounting for their change and their effects on individuals and other institutions. What sets him aside from a typical social historian or historical sociologist is his belief in the usefulness of formal game theory and models of rational choice in the study of particular institutions. What sets him aside from Guala is the historical and empirical orientation of his approach.

In sum, there is a stark contrast between Guala's philosophical high-level theorizing and Greif's applied analytic framework, which is comparative and context-specific. Whereas Guala's composite unification strategy promises field unification at an abstract level, Greif attempts to bring together insights from different social sciences under a common analytical framework to explain particular institutions. Guala aims in his functionalist account to foster understanding of the stability and persistence of institutions at a general level. Greif's framework, on the other hand, aims at explaining stability, change, and the effects of particular institutions—without assuming much about their origins and functions.

Lessons about unification and abstraction

We have briefly discussed three theories about social institutions with respect to three dimensions: unification, scope and explanatory goals. It has turned out that unification can mean many things. First, it could consist of a synthesis of existing theories. This composite unification brings together theories that have been developed separately to produce a more complex theory, which is what Guala does with game-theoretical ideas about institutions, incentivized action theory and Hindriks' reductive theory of constitutive rules. Similarly, he

synthesizes rule and equilibrium approaches to institutions. A second dimension of unification is field unification, which is also part of Guala's agenda. He aims to bring together social ontology and social-scientific approaches to social institutions. Here his approach is in sharp contrast with Searle's project to provide a general ontological theory about all social institutions. Searle's is theoretical unification of the domain via a comprehensive theory. Guala shares the same goal, but he has a more inclusive idea of the domain. For him, the general theory of institutions should bring together both philosophical and social-scientific theorizing. Underlying this ambition is a different conceptualization of the tasks of the general theory, and a different vision of the intellectual division of labor between the fields.

Turning to Greif, we find another set of ideas about unification. Both Guala and Greif pursue the goal of unifying the rule and equilibrium approaches to institutions, but they are doing quite different things. Whereas Guala aims at composite unification, Greif more modestly suggests that the two approaches are complementary. There is no reason to treat these conceptions of unification as mutually exclusive alternatives, however. Guala's notion is stronger: the successful achievement of his goal implies that Greif's goal is also achieved, but not vice versa. However, this observation holds only when we limit our attention to the unification of these two economic approaches to institutions. The reason for this is that, for Greif, the unification of the rule and equilibrium approaches is only part of a larger project. His aim is to produce a conceptual and analytical framework that also brings in ideas also from sociology, political science, cognitive science, and history. It is not at all clear that the completion of Guala's project would also complete Greif's more extensive version, which has many more elements to be integrated. Again, it should be noted that Guala and Greif operate on very different levels of abstraction: Guala aims at a general theory, whereas Greif's unification is achieved via a synthetic definition of institution that facilitates the incorporation of insights from different approaches into an analysis of particular institutions. What is missing from Greif's version is philosophical ambition: he is not intent on building a philosophical theory about institutions,

but is satisfied with enabling a multidisciplinary approach to facilitate their comparative and historical analysis.

	Philosophical reconstruction	Stability / Persistence	Change	Emergence
Searle	●			
Guala	●	●		
Greif		●	●	

Table 1. Explanatory goals

The diversity of theoretical aims among these three authors also carries implications for their explanatory goals. Despite its simplicity, Table 1 conveniently summarizes the differences and similarities between the three conceptions of the theory of institutions with respect to explanatory goals. Searle's only goal is to come up with an ontological reconstruction of the institutional reality. Such a reconstruction basically provides a philosophical explanation of the nature of institutions: it tells us what it is to be an institution. Greif, in contrast, aims at enabling explanations of stability and change in particular social institutions. The comparison between Searle and Greif makes it easy to understand why there has been little interaction between philosophy and the social sciences. Apart from the big differences in the levels of abstraction, the explanatory goals are quite separate. Guala, on the other hand, aspires to connect social ontology and the social sciences. He retains the goal of providing an ontological reconstruction of institutions, but he adds the goal of accounting for their stability and persistence, while explaining that the emergence and change of institutions remains outside his theory. Interestingly, this compromise is not arbitrary, but follows from the chosen level of abstraction: the goal of analyzing the generic features of institutions cannot accommodate the quite diverse elements of particular situations that are needed for the explanation of their emergence and change. None of the authors we discuss in this paper is interested in explaining the emergence of institutions. For this reason, the last column in Table 1 remains an empty placeholder for approaches such as evolutionary game theory and historical analysis, which at very different

levels of abstraction address the issue of how new institutions arise. It remains to be seen whether they can be incorporated into a "unified" theory of institutions.

The above discussion seems to imply that there are two separate projects in the literature. The first is the project of social ontology, the aim of which is to show how the social world fits into the world's ontological furniture. This kind of philosophical theory needs to account for institutions, language (whether or not it is regarded as an institution), and the diverse things Searle dismissively calls "systematic fallouts". We are still waiting for such a comprehensive theory, but it seems that we do have some idea of what it would look like. The second concerns the development of theoretical tools for analyzing and explaining institutions. As a social-scientific project it would purport to explain the emergence, stability, and change of real-world institutions. In contrast to the first project, it is not clear that such theoretical work will end up in a general theory of social institutions. Indeed, it is possible that Greif-style integrative theorizing might be more productive. What would be a proper level of abstraction for such theorizing is an open question. It may be that theorizing that is too general abstracts from the truly interesting stuff about the institutional reality.

This raises a concern about Guala's theoretical project: the preference for too abstract theorizing might be a bad equilibrium. Both social ontology and game theory address highly abstract and stylized *explananda*. Compared to other approaches in the social sciences, they are outliers in this respect. When these two are selected as the main components of theoretical synthesis, the result is also highly abstract. They focus mainly on "types" of institutions, whereas other approaches are primarily engaged in explaining particular institutions and understanding their relation to the historical and social context. We call this possible bad equilibrium, because it is possible that the abstract concerns of social ontology and game theory reinforce each other and make it easier to ignore the approaches that do not share their abstract focus. In other words, unifying the two might seem a significant achievement from the inside, but from the outside it

looks like entrenchment of the major obstacle to fruitful interaction between social ontology and social scientific analysis of institutions.

This basically challenges Guala's ability to cash his first promissory note. Let us consider sociology: although institutions are one of its main focuses of interest, there is very little general sociological theorizing about institutions that could be compared to social ontology and game theory. Why is this? It could be that sociologists are simply wrong, but it could also be that they have not found general theories very useful in terms of understanding real-world institutions. Their focus is very much on the emergence, change and effects of institutions. This is in sharp contrast with social ontology and game theory with their strong tendency to treat emergence and change as simply exogenous *explananda*, that can be dealt by giving some abstract pointers to salience/focal points, delta parameters, and actual motivations, for example. We do not deny the usefulness of this approach for certain purposes, but given its ambitions, Guala's theory should demonstrate its ability for fruitful engagement in empirical work on institutions; hence its fruitfulness and social-scientific legitimacy.

As our earlier discussion on the type-token distinction shows, the relation between general institutional concepts and particular institutions is more complex than Guala assumes. The rather simple toy models of institutions might be useful for the purposes of proof-of-the-concept theorizing, but extending these abstract results to the explanation of concrete institutional practices might be tricky. As Guala recognizes, particular institutions are historically changing, context-dependent things. For example, both Florentine and Mayan marriage are parts of a seamless web of other institutions and social practices, and as Guala is well aware, these institutions do not necessarily solve the same kind of coordination and cooperation problems. In fact, he argues that marriage solves several problems, and regulates a cluster of activities, none of which are essential but still constitute a natural kind (2016, 198–99). The basis for this claim is the assumption that "as long as a contract regulates (formally or informally) a cluster of

activities of this sort, it is legitimate to speak of marriage" (2016, 199). The key assumption here is that social institution types can be conceived of as social kinds that are analogical to natural kinds. According to Guala, "what marriage is, or what deserves to be called "marriage;' [...] depends on how the world is, in particular what kind of problems the institution of marriage (in its various instantiations) solves" (2016, 199). This is the most important (largely) implicit assumption in his approach, but although it is openly expressed in the last chapter, it is not explicitly defended in the book.

We remain skeptical about this natural-kind solution to the problem of general concepts in the social sciences. Perusal of the literature in the social sciences that study marriage (mainly anthropology and sociology) does not give any indication that they would finally reach consensus about the cluster of activities that define marriage. The same goes for property and money. It might well be that as far as economists are concerned, money has three central functions: as a medium of exchange, a unit of accounting, and a store of value. However, this is not the result of an extensive research of archaeological, anthropological, and historical sources, it is a theoretically motivated stipulation. We are not challenging the legitimacy of economists' conceptualization for many theoretical purposes. We are merely claiming that these functions do not define a natural kind that would serve as a type to all particular institutions of money.

In our view, the study of institutions should focus on particular institutions as historical causal complexes, not on their postulated abstract functional essences. From this alternative perspective, social mechanisms constitute the generalizable core of the social sciences, not natural-kind-based generalizations about institutions. Of course, there are many generalizations to be made about institutions. However, they are not universal but are rather bound to historical and institutional background conditions. Particular institutions such as Florentine and Mayan marriage are complex and historically changing. Because we conceive of institutions as historical individuals, we do not see any reason to assume that explaining their stability or persistence has

any explanatory priority. Explaining persistence is important, but so is accounting for the changes and origin. Analysts of institutions should aim at a causal account that will take care of both. This constitutes an alternative view of unification: whereas Guala is ready to accept that the rules-in-equilibrium account explains only the persistence of institutions and answers to other questions require a "different style of explanation" (2016, 31), our idea would be that a general causal account would accommodate all the relevant *explananda*. The motivation for our view is not purely philosophical. We also believe that the presumption of functional essences for institution types might make us blind to important changes in our institutional environment. For example, if one considers the future of money, it might be fruitful to relax the assumption that it always has three functions, or that it solves a specific set of problems (Birch 2017).

Conclusion

In this paper we have compared three very different visions of a unified theory of social institutions. We found out that unification can be many things, and that it may be associated with diverse explanatory goals. We have also presented some critical comments about Guala's project. However, our doubts should not obscure the bigger picture. We firmly believe that Guala's composite unification strategy is skillfully implemented and that field unification is a precious goal that is worth making every effort to pursue. Our aim was to facilitate the further pursuit of this goal by way of closely looking at Guala's theory in comparison to two prominent alternatives. What we have learned is simply the following: Guala's *Understanding Institutions* is an excellent first step, but there is still work to be done. Our suggestion to social ontologists is as follows: Follow Guala in taking seriously what the social sciences have to say about institutions, and then take it even more seriously.

References

Aoki, Masahiko. 2001. *Toward a Comparative Institutional Analysis*. Cambridge Mass.: The MIT Press.

Bates, Robert H., Avner Greif, Margaret Levi, and Jean-Laurent Rosenthal, eds. 1998. *Analytic Narratives*. Princeton University Press.

Birch, David. 2017. *Before Babylon, Beyond Bitcoin*. London: London Publishing Partnership.

Calvert, Randall R. 1995. "Rational Actors, Equilibrium and Social Institutions." In *Explaining Social Institutions*, edited by Jack Knight and Itai Sened, 57–94. Ann Arbor, MI: University of Michigan Press.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.

Greif, Avner, and Christopher Kingston. 2011. "Institutions: Rules or Equilibria?" In *Political Economy of Institutions, Democracy and Voting*, edited by Avner Greif and Christopher Kingston, 13–43. Dordrecht: Springer. https://doi.org/10.1007/978-3-642-19519-8_2.

Guala, Francesco. 2016. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton: Princeton University Press.

Guala, Francesco, and Frank Hindriks. 2015. "A Unified Social Ontology." *Philosophical Quarterly* 65 (259): 177–201. https://doi.org/10.1093/pq/pqu072.

Hindriks, Frank. 2005. "Rules and Institutions: Essays on Meaning, Speech Acts and Social Ontology." Erasmus University Rotterdam.

———. 2009. "Constitutive Rules, Language, and Ontology." *Erkenntnis* 71 (2): 253–75. https://doi.org/10.1007/s10670-009-9178-6.

Hindriks, Frank, and Francesco Guala. 2014. "Institutions, Rules, and Equilibria: A Unified Theory." *Journal of Institutional Economics* 11 (3): 459-480. https://doi.org/10.1017/S1744137414000496.

Lewis, David. 1969. *Convention*. Cambridge: Harvard: Harvard University Press.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Schotter, Andrew. 1981. *The Economic Theory of Social Institutions*. New York: Cambridge University Press.

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

———. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.

———. 2015. "Status Functions and Institutional Facts: Reply to Hindriks and Guala." *Journal of Institutional Economics* 11 (03): 507–14. https://doi.org/10.1017/S1744137414000629.

Smit, J. P., and Filip Buekens. 2017. "How to Do Things Without Words: A Theory of Declarations." *Philosophy of the Social Sciences* 47 (3): 235–54. https://doi.org/10.1177/0048393117697073.

Smit, J. P., Filip Buekens, and Stan du Plessis. 2011. "What Is Money? An Alternative to Searle's Institutional Facts." *Economics and Philosophy* 27 (1): 1–22. https://doi.org/10.1017/S0266267110000441.

———. 2014. "Developing the Incentivized Action View of Institutional Reality." *Synthese* 191

(8): 1813–30. https://doi.org/10.1007/s11229-013-0370-5.

Sperber, Dan. 1996. *Explaining Culture. A Naturalistic Approach*. Oxford: Blackwell
Publishers.

Sugden, Robert. 1986. *The Economics of Rights, Co-Operation and Welfare*. *The Economics of Rights, Co-Operation and Welfare*. New York: Palgrave Macmillan.
https://doi.org/10.1057/9780230536791.

Biographies

N. Emrah Aydinonat is a senior researcher at TINT (Centre for Philosophy of Social Science) at the University of Helsinki. His research focuses on the epistemic value of idealized models in economics and theories of social institutions. He is the author of *The Invisible Hand in Economics* (Routledge, 2008) and the co-editor (w/ J. Vromen) of *Economics Made Fun: Philosophy of the pop-economics* (Routledge, 2015).

Petri Ylikoski is Professor of Science and Technology Studies at the University of Helsinki. His research interests include theories of explanation and evidence, science studies, and social theory. His current research focuses on the integration of findings from biological sciences (neurosciences, genetics, and evolutionary biology) into the social sciences and the foundations of mechanism-based social science.