Against Guala and Hindriks' Functionalist Theory of Institutions

Louis Larue

Abstract: What explains the existence and persistence of institutions? This article centres on Guala and Hindriks' functionalist theory of institutions, which explains their existence and persistence by their overall beneficial consequences, where these consequences are not the intentional product of individual or collective human decisions. According to them, institutions exist and persist because they generate "cooperative benefits," through their ability to solve coordination problems. This article aims to show that their theory is lacking in at least three respects. First, indeterminacy in the selection of coordination devices weakens its predictive power. Second, their account relies on an inappropriate conception of the benefits of institutions, which makes it unable to explain why some institutions exist and persist while others do not. Finally, it lacks empirical support from history.

Keywords: Functionalist explanations, Institutions, Game Theory, Functions, Status Functions

1. INTRODUCTION

Explaining the nature of institutions, and institutional facts, has been the purpose of many economic and philosophical theories. Broadly speaking, one

10.25365/jso-2025-9106
Published online November 05, 2025

Louis Larue, Aalborg University, Denmark, E-mail: louisl@ikl.aau.dk

might delineate two opposite views on the matter. On the one side, economists such as North (1990, 6) and philosophers such as Searle (1995; 2010) and Tuomela (2002; 2007) argue that institutions are rules that guide behaviour. According to Searle, for instance, institutional facts are created by the collective assignment of specific functions upon pre-existing objects, persons or state of affairs. Thus, a piece of paper is money in a certain context because people collectively recognise it as money, and impose a status upon it, which in turn enables that piece of paper to perform certain functions (means of payment, etc.). On the other side, following Lewis' ground-breaking account of conventions (Lewis, 1969), game theorists argue that they are solutions to repeated coordination problems which arise from people's strategic interactions (Bicchieri, 2006; Binmore, 2010; Hédoin, 2017; Sugden, 1986). As an example, one may explain the fact that, in Sweden, one must drive on the right side of the road as the equilibrium outcome of a repeated coordination game. Some authors, such as Greif and Kingston, and Guala and Hindriks, have also proposed theories which attempt to bridge the two views by accounting for rules within game-theoretic frameworks (Greif and Kingston, 2011; Hindriks and Guala, 2015).

Within this literature, some authors give great importance to the explanation of the existence and persistence of institutions.[1] For instance, North (1990) proposes a complex account of how a certain pattern of bargaining between political and economic actors explains the emergence of specific property right regimes in the Western world. He then argues that interests' lock-ins sometimes hinder institutional change and prevent societies from reaching efficient institutional arrangements. This article will not review all the theories purporting to explain the existence and persistence of institutions. Its focus will be on functionalist theories of institutions, according to which institutions exist and persist because they fulfil a function. More precisely, functionalist theories explain the existence and persistence of institutions by their overall beneficial consequences, where these consequences are not the intentional product of individual or collective human decisions.

This latter distinction is important. Searle (1995), for instance, argues that functions are crucial to the understanding of institutions, but that they are assigned by humans intentionally and collectively (through what he calls "collective intentionality"). For this reason, I do not count his theory as part of the functionalist camp.

1 Not all. Searle, for instance, explicitly disregards these questions (Searle 2015, 507–14).

The clearest defence of a functionalist theory of the existence and persistence of institutions has been given by Hindriks and Guala (2021). According to them, institutions exist because they generate "cooperative benefits." They write that "generating cooperative benefits is the causal or etiological function of institutions because it serves to explain why institutions exist and persist" (Hindriks and Guala 2021, 2028).

Their theory is amongst the most influential and developed philosophical attempts to defend a functionalist theory of institutions in social ontology. It is also perhaps amongst the most controversial. This article will focus on criticizing the theory developed by these philosophers. Section 2 presents their approach. Section 3 tries to clarify some of the key concepts of their view, namely the difference between existence and persistence, and between theories focusing on specific institutions and theories focusing on institutions in general. Section 4 focuses on the part of their theory dealing with the existence of institutions while Section 5 centres on their explanation of the persistence of institutions. Section 6 summarises the main conclusions of the article. First, I argue that their account lacks predictive force. Second, I show that their account relies on a conception of the benefits of institutions that makes it unable to explain why some institutions exist and persist, while others fade out. Finally, I contend that Guala and Hindriks unjustifiably disregard historical and empirical facts.

Though this article concludes that Hindriks and Guala's account faces serious difficulties, it must be noted that this does not entail that all functionalist explanations of the emergence and/or persistence of institutions are generally doomed to fail. Proving the latter claim is far beyond the reach of this article and I doubt that it can be done. But I believe that my objections against Hindriks and Guala's account may serve as a reminder that building rigorous functionalist explanations in social ontology faces important challenges and that these challenges cannot be ignored.

2. GUALA AND HINDRIKS' FUNCTIONALIST ACCOUNT OF INSTITUTIONS

It is an uncontroversial fact that institutions have functions—good or bad, effective or not. It is also a platitude that all institutions bring about some benefits to some people at least sometimes. What is more controversial is the role that functions play within theories of institutions, as well as how they arise.

A first issue of contention concerns how institutions acquire their functions. How does money acquire its capacity to serve as a means of exchange,

for instance? And how does a river become a border?

For some, institutions acquire their functions through collective imposition: people collectively intend a specific piece of paper to serve as a means of exchange and a specific river to serve as a border. In that sense, collective intentionality is crucial to Searle's and Tuomela's account of the function of institutions (Searle 1995; 2010; Tuomela 2002; 2007). Searle assigns great importance to collective intentionality in his account of institutions. He writes repeatedly that "institutions are collectively accepted systems of rules" and speaks of the "collective assignment of functions" (Searle 1995, 21–2). Searle's insistence on collective intentions may seem strange, for we generally think of intentions as a feature of individual agents. How can we make sense of "we-intentions"? Let me start with an example, taken from Butchard and D'Amico (2015). When we say that two people are walking together in the same direction, we might mean, first, that we observe that they both happen to walk side by side, each intending to go in that direction, but with no collective intention of doing so. They are "alone together," as Butchard and D'Amico nicely phrase it. This shows, according to these authors, that joint action is not a sufficient condition for collective intentionality. Rather, when we say that two persons are walking together as part of a collective intention, we mean that they each take part in the joint action and share the joint intention of walking together. "My" doing this action is part of "our" doing this action.

Hindriks and Guala (2015), Hédoin (2017), Smit, Buekens and du Plessis (2011; 2016), among many others, strongly oppose the existence of collective intentions, or, at least, they argue that they are not necessary to explain how institutions acquire their functions. They contend that game-theory is able to explain the existence of functions by relying only on individual actions and individual intentions. For instance, the rule "drive on the right side of the road" may be explained as the equilibrium outcome of a game without collective coordination or intention.[2]

As this article centres on Hindrik's and Guala's theory, I will mainly analyse their own attempt, which is part of a larger literature whose main effort is to build game-theoretic accounts of institutions. That literature started out with Lewis's study of conventions (Lewis, 1969), and has had a fruitful life of its own in economics (For a review, see Binmore 2010).

2 For a more detailed comparison of different viewpoints on collective intentions, see Larue (2024, 721–41).

Guala and Hindriks argue that institutions are solutions to coordination problems. More specifically, they claim that institutions are "correlated equilibria of coordination games with multiple equilibria" (Hindriks and Guala 2015, 466; see also Guala and Hindriks 2015, 182–6). A coordination game is a game in which several agents (say two) must decide what action to undertake (their "strategy"), and in which the outcome for each player depends on the other player's action. An equilibrium of a coordination game is a profile of strategies (one for each player), where each player's strategy is the best response to those of other players. Guala and Hindriks take the example of two tribes having to decide where to hunt. There are two possible hunting grounds, and both tribes cannot hunt at the same time in the same place. There are several equilibria in this game, for they can each hunt in either hunting ground. How will they agree on which land to hunt? (Suppose that they do not talk to each other). One way to solve the problem is to use correlation devices, that is, arbitrary "signs" or "pre-emption devices," such as "whoever happens to be there first hunts first." In technical terms, we say that a correlated equilibrium involves strategies that are conditional upon an event or signal sent by an external coordination device. For Hindriks and Guala, these coordination devices are "rules." This is why they say that their account is a unification of the rules and equilibria approaches to institutions (Guala, 2016; Guala and Hindriks, 2015).

Therefore, their account does not assume collective intentions. It is "collective" only in so far as people interact with each other within the coordination game. But people's intentions are strictly individual. Nor does their account presuppose any kind of collective imposition of functions. According to Guala and Hindriks, institutions can be modelled as the unintended solutions of coordination games. According to their model, for instance, the fact that we drive on the right side of the road in many countries may be explained as a convenient yet unintended solution to a common coordination problem: it might be that people just started to walk on the right side of roads in these countries, and that became the norm—without any collective intention to do so. As a corollary, the fact that this norm ('drive on the right side of the road") functions as a coordination device and thus solves a coordination problem, has not been imposed either on any object, person or state-of-affair by an individual or a group. Rather it is the unintended and beneficial consequence of the emergence of institutions.

To be clear, each player does have individual intentions, which relate to maximizing their own advantages (or utility) within the game. Yet no player has the intention—whether collective or individual—to impose a function

on something in order to solve the game. The point of Hindriks and Guala's game-theoretic framework is that institutions can be modelled as unintended equilibrium states which result from strategic individual interactions. In game theory, no individual, and a fortiori no collective, ever intends to reach a specific equilibrium. They only look at their own advantage, and choose their strategy as a function of the strategies that others might choose. Hence, equilibria are the ex-post products of strategic interactions, in which all players' strategies are best responses to each other's strategies.

At this stage, Guala and Hindriks go a step further and claim, in addition, that the function of institutions (solving coordination problems and thus producing cooperative benefits) also explains their existence and persistence (Hindriks and Guala 2021, sec. 3.2). In more technical terms, they write that it is the "etiological" function of institutions to provide those cooperative benefits, because these cooperative benefits explain their existence and persistence (Hindriks and Guala 2021, 2028). Hence, their theory allows to model the existence and persistence of institutions by reference to strategic interactions between actors, rather than as the intended product of the collective or individual imposition of functions.

Because of these features, their theory is undeniably functionalist: it relies on the unintended beneficial consequences of a phenomenon to explain the existence and persistence of this phenomenon.

The use of etiological functions and of functionalist explanations is common in evolutionary biology, which explains the evolution of species in part by reference to the beneficial consequences, for a given species' reproductive success, of unintended genetic transformations. As far as these consequences explain evolution, they perform an etiological function. As we shall see in the next section, however, the explanatory role of functions is more controversial in other scientific fields (Elster, 1983). Searle, for instance, denies that they have a role to play for theories of institutions and strongly opposes functionalist explanations of the existence of institutions (Searle 1995, 14–20).

3. CONCEPTUAL CLARIFICATIONS

As a first step in my critique of Guala and Hindriks' theory, I would like to make two distinctions.

First, a theory can focus on specific institutions or on institutions in general. For instance, one may aim to explain why we have this specific private property regime in Sweden or why the border between Sweden and Denmark lies where it now lies. Or we can aim to explain why we have property rights

and borders in general, without any specific example in mind. Yet, in my view, one can ask questions about institutions in general only if there also exist corresponding specific institutions. As I will argue below, in order to answer questions on institutions in general (for instance, on their existence and persistence), one cannot abstract entirely from specific examples. One needs these examples to show that one's theories and explanations are not mere plausible stories, but that they in fact have real-world applications on which one can test one's claims.

Second, there is a distinction between the existence and the persistence of institutions. For Hindriks and Guala (2021, 2032), persistence means "continued existence." But what does existence mean? They are less clear on that matter. It must mean something different than persistence. For they themselves make this distinction and always speak of the "existence and persistence" of institutions. I think that the most plausible understanding of existence in their work is that it is equivalent to emergence.

Hindriks and Guala have borrowed and adapted the concept of "etiological" function from Wright, who defined it in relation to functional explanation. Wright (1973, 156) writes that "functional explanations, although plainly not causal in the usual, restricted sense, do concern how the thing with the function *got there*. Hence, they are etiological, which is to say "causal" in an extended sense" (Wright's own emphasis). For Wright, the etiological functions of a phenomenon are "causal" in the sense that they explain "how we got there," that is, its emergence. If Hindriks and Guala's understanding of etiological functions is in line with Wright's, then, they must mean that explaining the existence of an institution by reference to its etiological functions amounts to explaining how it came about, how we got there, that is, it amounts to explaining the emergence of institutions.

Apart from referring to Wright, Hindriks and Guala also write that the etiological function of an entity has two roles in their own theory. First, it can explain why that entity "continues to exist (or, in other words, why it persists)" (Hindriks and Guala 2021, 2032); second, they consider the explanation of its existence and write that:

> Equilibrium accounts are usually less informative when it comes to explaining how an entity comes into existence. We do not know of a satisfactory and broadly applicable account of the emergence of institutions. Note, however, that the attribution of a function to the heart does not require a complete history of this organ. What is required instead is a convincing account of how the fact that the entity does F

> explains its existence. It is rather plausible that the fact that institutions generate cooperative benefits fulfils this role. Arguably, these benefits play a crucial role in the selection process that only some institutions survive. (Hindriks and Guala 2021, 2032–3)

This passage clearly indicates that, for Guala and Hindriks, existence means emergence and that explaining the existence of an entity amounts to explaining how it "comes into existence." Hence, from now on, existence will be taken as a synonym of emergence. But this paragraph also illustrates the ambivalence of these authors regarding the ability of their account to explain the existence of institutions. On the one hand, they seem hesitant on the potential of success of equilibrium accounts on that regard. In a footnote appended to this paragraph, they even acknowledge that "we only account for the tight link between attributions of functions and explanations of the continued existence of a type of entity, while noticing that its emergence is often a haphazard process." (Hindriks and Guala 2021, 2033, n. 8). On the other hand, however, they write that "a convincing account" may suffice for explaining the emergence of institutions and that their own theory provides such a "plausible" satisfactory explanation. Hence, they seem to acknowledge that, despite the failure or lack of appeal of previous attempts at explaining the emergence of institutions, their own theory will not be equally unsuccessful. Finally, I wonder why they would constantly insist that their functionalist account aims to explain the persistence *and existence* of institutions if they believed that the latter task was doomed to failure. Hence it is reasonable to interpret their theory as aiming to explain the emergence and the persistence of institutions.

The next two sections will inquire whether the doubts of Guala and Hindriks are justified: can their own equilibrium account of institutions explain the emergence and persistence of institutions by appealing to their cooperative benefits?

4. EMERGENCE

4.1. Emergence and Indeterminacy

In this section, I criticise the predictive potential of their theory of the emergence of institutions by pointing out that it is unable to predict a priori which coordination device will in fact emerge.

Take the following coordination problem: On which side of the road should we drive? For Guala and Hindriks, that question is actually equivalent to asking: What coordination device could solve the issue? Here are possible

answers:

- Always drive on the right
- Always drive on the left
- Toss a coin each time you meet someone to decide on which side to drive
- Drive on the right on weekdays, and on the left otherwise
- …

As we see here, there is a large number of possible solutions to this problem. How can we select among them? Common sense would favour one of the two first coordination devices. But not Guala and Hindriks' account.

Note, first, that the selected coordination device need not be a "good" device. All it needs to do is to solve the coordination problem at hand, something that all the devices listed above are able to do. According to their account, the quality of coordination devices only refers to their ability to solve a coordination problem. In that sense, all coordination devices, if they are able to solve the coordination problem, are equally beneficial (see section 4.2 below for a more detailed discussion of that point). Their account is thus silent concerning the efficiency, equity, or legitimacy of such devices. It can happen of course that, in the example discussed above, drivers are lucky and discover the most efficient and equitable arrangement just by chance, but there is no necessity for this in Guala and Hindriks' account.

It is worth noting that allowing their account to make predictions on the efficiency, equity or legitimacy of institutions would either break their promise to reject intentionality, or distort their own understanding of the benefits of institutions. Both authors clearly reject intentionality in the selection of coordination devices, as explained in section 2. They also make clear that the etiological function of institutions is limited to their ability to solve coordination problems. I think these two claims hold tightly together. Adding other benefits would necessarily break the claim that there is no intentionality in their account. For how can the chosen coordination device be the most efficient, or even adequately efficient, if its discovery is purely accidental and not driven by the intentions of the actors involved?

This may not be a serious problem for their account. For its goal is to explain the emergence and persistence of institutions, not their efficiency, equity or legitimacy (or lack thereof). However, it limits its scope. Because their account restricts the definition of the benefits of institutions to their ability to solve coordination problems, and because it leaves no place to intentionality

in the choice of coordination devices, it must remain silent on the efficiency, equity, and legitimacy, of coordination devices and of institutions.

The "silence" of their account on efficiency is of some consequence for the second, and more important, point that I want to discuss, and which concerns the selection of the coordination device that will, in practice, solve the coordination problem. As the example above shows, for every coordination problem, there is, potentially, a large number of adequate coordination devices available. Yet their account is silent concerning which of these devices will, or has to be, selected a priori: strictly speaking, any device that is able to solve the coordination problem at hand could emerge as a possible solution. Moreover, as I just showed in the previous paragraphs, their theory cannot accommodate the (plausible) conclusion that, in practice, the most convenient solution will be chosen. The indeterminacy of their theory is a problem because it severely restricts the predictive power of their account.

As we shall see in the next section, their account is likely to be more successful as an explanation of the persistence of existing institutions. In that case, it can be pointed out that already existing institutions persist because they help to solve coordination problems. But if Guala and Hindriks' purpose is to explain the emergence of institutions, their account will run the risk of indeterminacy.

For instance, their account is unable to explain why, in Sweden, people drive on the right, while, in the UK, they drive on the left, other than by reference to chance. It could in fact have predicted that people would have driven on the right in both countries, or on the left, or any other arrangements providing a solution to this peculiar coordination problem.

One way, perhaps, to improve their theory is to think of the emergence of institutions as a slow process of selection. Once people begin to travel, they will meet other people on the roads. If they interact as described by Guala and Hindriks, each time they meet, they will face a coordination problem: on which side of the road should we drive (or walk)? We can think that the first times, unknown to them (since we are in a social world without collective intentions), some coordination device arrived at randomly will help them. For instance, because most people are right-handed and must be able to hold their sword in case of attack, they walk on the left side of the road.[3] In other instances, other people with other characteristics may follow other rules. But, as the number of encounters increases, one of these rules will prevail, at least within some

3 For other possible hypothesis, based on a large amount of sources from neurology, history and anthropology, see McManus (2002). The author, though, makes clear that many of these hypotheses involve a significant amount of speculation.

geographic zone of exchange, because people just keep on doing what they and others did in previous occasions, and because they may have a natural tendency (again, unintentional and perhaps even unknown to them) to disregard the more cumbersome rules.[4] Hence, the minority rules will soon disappear, and one institution will emerge victorious. In short, what I have sketched here is a slow process of selection.

Is this a credible story? Let me say, first, that it is a possible story, though not one explicitly defended by Guala and Hindriks.[5] The credibility of this story depends on the credibility of the assumption that people do not decide intentionally on the rule they will follow, but that the rule is decided instead by the presence of a coordination device (such as the fact that most people are right-handed and need to bear a sword). One could argue for another credible story, one in which people would actually talk and deliberate with each other and decide, perhaps informally, perhaps at an official gathering, on which side of the road people should be walk. Coordination devices may still have a place in this alternative story, for people may consciously decide to walk on the right side, for instance, because they think that actions performed on the right side have a special religious or political significance unrelated to the act of traveling.[6]

Which of these stories is the right one? Contrary to what Guala and Hindriks claim, we cannot decide on the credibility of a story just based on its prima facie plausibility, for different, alternative stories may be equally plausible. In the next subsections, I will try to argue that the credibility of stories of that sort, in the last instance, depends on the empirical data that we have at our disposal, that is, on the study of history.[7]

To sum up: for Guala and Hindriks, just any possible device that can solve a coordination problem may constitute an adequate coordination

4 I confess that this sentence is a bit obscure. However, the constraints set by their account leaves no place to a more precise description of how one rule finally prevails. Note, moreover, that contrary to the theory of natural selection in evolutionary biology, their own theory cannot refer to the ex-post benefits of the selected coordination device to explain its comparative success, because, according to their own conception of the benefits of institutions, all coordination devices, and hence all institutions, have the same benefits, namely, solving coordination problems. No coordination device is comparatively better than another. See section 4.2 below for an elaboration of this point.

5 Though they hint at this possibility when they write that the cooperative benefits of institutions "play a crucial role in the selection process that only some institutions survive" (Hindriks and Guala 2021, 2032).

6 See McManus (2002) for a discussion of the religious and social significance of left and right.

7 For instance, see Poehler (2017) for an interesting account of traffic rules in the Roman empire.

device. I argued that the indeterminacy of the selection of the device weakens the predictive potential of their theory. Because there are a lot of potential candidates to the role of "coordination device," their account is unable to explain the emergence of specific institutions (for instance, the fact that we drive on the right).

4.2. Emergence and the Benefits of Institutions

A second worry concerns the alleged "benefits" of institutions, which, Guala and Hindriks argue, relate to their cooperative benefits, or the fact that they solve coordination problems.

A good theory should explain why some institutions emerged while others, which could have emerged, did not. Hindriks and Guala argue that what explains the emergence of institutions is the fact that they solve cooperative problems (i.e. their etiological function). The issue is that for every given coordination problem, there is several (if not many) potential coordination devices that are able to solve it. Take the driving example again. As I have shown above, there are plenty of equilibria in this game, and each would count as an institution according to Guala and Hindriks' account. By definition, each of these equilibria would be a solution to a coordination problem, and thus constitute an institution generating cooperative benefits. Hence, the mere fact that institutions provide cooperative benefits cannot explain why we have the institutions that we have now, and not others, equally beneficial. For, in the restricted sense given by Hindriks and Guala, all institutions, whether actual or only possible, are beneficial.

Hindriks and Guala's account is thus unable to explain why specific institutions arise while others do not, for all possible institutions do, by definition, solve coordination problems and are thus possible candidates. They could all have emerged and persisted over time. If that is what these authors have in mind, the fact that institutions produce cooperative benefits is both a very weak understanding of "benefit" but also, and more fundamentally, a very inadequate explanatory tool for a theory of the emergence of institutions. Not only is the "choice" of the coordination device indeterminate, but all coordination devices are by definition beneficial. Because of these two features, their theory cannot make predictions on the emergence of specific institutions.

4.3. History and the Need for Empirical Evidence

The natural way to go when studying the emergence of institutions would be to look at history. Unfortunately, Hindriks and Guala do not provide any empirical evidence for their theory of the emergence of institutions. Instead, they seem to give preference to the possibility of reconstructing rationally how institutions could have emerged in a game-theoretic setting. For instance, when discussing functionalist explanations in biology, they write that "the attribution of a function to the heart does not require a complete history of this organ. What is required instead is a convincing account of how the fact that the entity does F explains its existence" (Hindriks and Guala 2021, 2033). I doubt, however, that a plausible theoretical account may suffice.

One example of a successful functionalist theory may be found in evolutionary biology, which explains, at least partly, the evolution of species by reference to the reproductive benefits of certain random genetic variations (Kitcher, 2009). But evolutionary biologists and geneticists have collected an impressive amount of evidence in support of their theories—and none, to my knowledge, would claim that all that is needed to explain evolution is a fitting theory. Hindriks and Guala claim that "the attribution of a function to the heart does not require a complete history of this organ" (Hindriks and Guala 2021, 2033). This is certainly true if the purpose of the inquiry is to explain how the function of that organ explains its persistence. But explaining the emergence of that organ does require some look at history, as does explaining the emergence of institutions. I agree that the history of a specific organ or institution should perhaps not be complete, but it should not be entirely disregarded either. Studying the history of the evolution of living beings is precisely what evolutionary biology has been doing at least since Darwin. Of course, different fields of science might follow different methodological rules, but the success of evolutionary biology is in part explained by its great care in collecting empirical evidence, and I do not see why functionalists in institutional theory could not follow its example.

In support to their preference for a rational reconstruction of the emergence of institutions, Hindriks and Guala also write that they "do not know of a satisfactory and broadly applicable account of the emergence of institutions" (Hindriks and Guala 2021, 2032–3). I agree. In fact, I do not myself know of a convincing theory of the emergence of institutions. However, the fact that we do not presently have a convincing account of the emergence of institutions that is supported by adequate empirical evidence does not entail that attempts to reconstruct their emergence theoretically are to be

preferred. There is in fact a large literature on the historical emergence of institutions. In economics, for instance, North has won the Nobel Prize for his account of the nature and emergence of institutions (North, 1990).[8] In anthropology, Graeber (2011) has offered his own historical account of the emergence of money and debt. In sociology, Weber's theory of the rise of capitalism is still influential (Weber, 1904), while historians, such as Braudel (1988), have studied in depth the emergence of market institutions. None of these accounts may be right, of course. And they do not all share the same aims and the same methodological commitments than Guala and Hindriks's account. But they demonstrate at least that an inquiry into the history of the emergence of institutions is possible. If Guala and Hindriks believe that these attempts are unsuccessful, they owe us an argument for why this is the case and why rational reconstructions are to be preferred.

More fundamentally, I want to argue that even if it were the case that Hindriks and Guala's account provided a theoretically convincing functionalist explanation of the emergence of institutions, this would be no proof that actual institutions are really the product of functionalist game-theoretic processes. It is certainly possible to build up a theoretical story of how money, debt, markets, and other institutions, have emerged through such processes. However, without proper use of at least some empirical evidence, these are just stories (on this point, see Elster 2015). To be clear, I think that building "purely" theoretical explanations of a phenomenon without prior consideration for empirical facts is a valuable scientific endeavour, at least as a first step. My point is that their theory needs to go through some sort of empirical test for it to be convincing or credible.

The need for empirical evidence is particularly crucial for functionalist explanations, such as Guala and Hindriks' functionalist theory of the emergence of institutions. As Elster (1983) has convincingly argued, successful functionalist explanations need a feedback loop in order to demonstrate how consequences (which happen logically and temporarily after the phenomenon they are supposed to explain) can "cause" an event (such as the emergence of a specific institution). Indeed, one may wonder how the cooperative benefits produced by an institution can explain the emergence of that very institution, given the fact that these benefits happen after its emergence.

Without evidence of the existence of a feedback loop, the undeniable fact that institutions produce benefits is no proof that these beneficial consequences explain their emergence. Plenty of institutions do serve well certain useful

8 His work has even led to an entire school of thought, see Greif (2006).

purposes. Since the history of their emergence is often obscure, it may well be tempting to claim that these very useful purposes are in fact what explains their existence. Yet, without some empirical evidence showing how ex-post consequences actually can become "causes," these explanations are just unwarranted.

In evolutionary biology, natural selection plays the role of the feedback loop. Certain random genetic transformations are beneficial to the reproductive success of a given species in a given context and thus feeds back into the explanation of its observed evolution. But what is the feedback loop for functionalist theories of institutions?

First, one would be tempted to look for existing examples in the history of human institutions—a step, though, which is absent from Guala and Hindriks' account. Yet, it is hard to come up with a convincing real-world example because it is hard to observe the selection process at work for existing institutions. We would need to observe how different competing institutions have fared differently depending on how their characteristics fitted their environment. This is a very difficult empirical task. Biology has millions of years of evolution to study, while we have a few decades or at best centuries of institutional change. Moreover, as many economists and political scientists have shown, plenty of inefficient and inadequate institutions survive over time, for various reasons (Elster, 1983; North, 1990).

Another possibility for Guala and Hindriks is to come up with a theoretical story. One could view the theoretical history of institutions as a series of reinforcements: after it has emerged, the existence of a given institution is reinforced by the incentives it gives to its users to stick to it. As in the story I told in section 4.1, once we have an institution stating that one should drive on the left side of the road, people have an incentive to do so if they want to avoid collision. This institution is reinforced every time two persons meet and abide by the rules of the institutions ("drive on the left if the other is also doing so"). In that case, the feedback loop is the fact that, once the rule has emerged, everyone has an incentive to avoid the negative consequences attached to deviation from it.

I do think that incentives could play the role of the feed-back loop and that theoretical constructions of the emergence of institutions, such as Guala and Hindriks', remain interesting to make sense of historic examples of institutions. But only the study of history can tell us what the origin of this specific institutions is, for the question of the emergence of specific institutions is a historical question (Aydinonat and Ylikoski 2018, 564). Without some empirical evidence for their account, their story will remain a mere plausible

story.

5. PERSISTENCE

Why do institutions persist? Why do people stick to an institution? Why do they obey rules? Hindriks and Guala argue that the failure to explain why people stick to certain institutions is a central flaw of all rules-based accounts of institutions (Hindriks and Guala 2015, 462). They defend their game-theoretic account of institutions partly on the basis that it can explain the persistence of institutions. They use two distinct arguments. One has already been presented above: institutions persist because of their etiological function, i.e. their capacity to solve coordination problems and thus to produce cooperative benefits. In a series of other papers, they also argue that their account has the additional advantage of securing some place to incentives, and thus can explain why certain institutions sometimes persist and why some institutions sometimes fail.[9] For instance, people keep driving on the right side of the road in countries in which this has become the rule because this is the best response to what others are doing. Individually deviating from the rule would lead to disastrous consequences for individuals, who thus have an incentive to stick to the rule.

In the following sections, I will focus on each argument in turn.

5.1. Persistence and the Benefits of Institutions

A first worry concerns the alleged "etiological benefits" of institutions. For Hindriks and Guala, the persistence of institutions may be explained by their *beneficial* functions (not simply their functions) which, they argue, relate to their cooperative benefits, or the fact that they solve coordination problems. I have already discussed the problems induced by this view when dealing with the explanation of the emergence of institutions. I would like now to focus on the explanation of their persistence.

A good theory should explain why some institutions persist while others die out. If Guala and Hindriks are right, beneficial institutions (in their specific sense) should survive, while non-beneficial ones should fade out. At this point, many readers may point out that plenty of failed institutions do persist over very long periods of time. North (1990), for instance, studied in detail the case of property right regimes in Latin America, which, according to him, are

9 Other authors have had recourse to a similar strategy. See, for instance, Smit et al. (2011; 2016).

deeply inefficient and lead to consistently lower rates of economic growth than those that could be expected. Moreover, as Eriksson (2019) points out, even if an institution benefits some people at one point in time, its benefits may fade out over time or end up being advantageous to only a fraction of society. Finally, not only inefficient but also morally bad institutions, such as slavery, also persist over time.

However, even such undesirable institutions do provide cooperative benefits. Bad legal regimes are better than no legal regimes at all because they help solving coordination problems, which otherwise would plague us. More controversially, immoral legal regimes such as slavery also produce cooperative benefits—by solving coordination problems—and may thus be seen as "beneficial" in Guala and Hindriks' specific sense.

The greater problem with Guala and Hindriks' conception of the benefits of institutions is that it is far too general to be able to explain why some specific institutions persist while others do not. To repeat, *all* institutions can be said to provide cooperative benefits. Why would some institutions fade out, then? Why do some institutions flourish and not others? If all institutions are (by definition) beneficial and if their benefits explain their persistence, no institution should ever change or die out.

Hindriks and Guala may reply that they are instead aiming at explaining the persistence of institutions in general, and not of specific institutions (even if they do not seem to consider this distinction). For instance, their framework may aim to explain why we have laws, and not why we have this specific legal regime in this country at that time. In that broader sense, they could claim that property laws (whatever their exact content) persist because, without them, we would be stuck in endless disputes on ownership, rights, and other property-related issues, at every single moment of our lives.

I think this move would make their theory uninteresting. We want to know why capitalist markets have persisted over time, and not other forms of economic arrangements; or why we have the laws that we have and not other laws; etc. Their theory holds that all institutions, once they have emerged, will produce cooperative benefits, and will thus be able to persist. Unfortunately, their theory is thus unable to explain why specific institutions persist and others fade out.

5.2. Persistence and Incentives

Can incentives explain the persistence of institutions, if cooperative benefits cannot? In order to answer this question, let me come back to Guala and

Hindriks' discussion of intentionality, whether collective or individual.

The rejection of intentionality from the explanation of the emergence and persistence of institutions is a crucial component of Guala and Hindriks' account. First, they explicitly argue that collective intentionality, or the view that we can have we-intentions distinct from I-intentions, is not necessary to a successful account of the persistence of institutions, and can thus be discarded. Second, they also argue that individual intentionality amounts to no more than the intentional pursuit of one's best interests within strategic interactions. People do not have the intention to create institutions with others, even if it may be to their own benefit to do so (see section 2). However, once institutions exist, Guala and Hindriks argue that people stick to them because it is in their best interest to do so given the incentives produced by these institutions.

Once again, their argument is ahistorical. They argue that, since it is possible to construct a game-theoretic model explaining the persistence of institutions by reference to some institutional incentive structure, it is thus not necessary to appeal to notions such as collective intentionality, whose exact nature may be hard to grasp. My objection to their view is that, even if such a theoretical model is possible, and even if they are right that collective intentionality or other obscure notions are superfluous, this does not suffice to show that incentives are what explains persistence. As in the case of emergence, their explanation of emergence must have some consideration for empirical evidence.

It is undeniable that people often have incentives to obey institutional rules (such as driving on the right) since deviating from the rule will be costly for them. However, are these incentives really what explains the persistence of institutions? Many have doubts about this. Searle, for instance, argues that sometimes the incentive to follow a rule may disappear and yet people continue to act according to the rules (Searle 2015, 511–2). Some people keep their promises even if they no longer have an interest in doing so, people pay their taxes even in the absence of fines, drivers respect the speed limits even in the absence of the police or of any other driver on the road, etc. Hence, Searle claims that institutions give people desire-independent reasons for action.

No theoretical argument can arbitrate this debate. Though many of Searle's arguments are sound, he does not seek any empirical support for them. If, like Searle, we want to reject the view that incentives are the driving force behind the persistence of institutions, if we want to argue instead that desire-independent reasons for actions, such as duties or customs, explain why people stick to them, what we need is some empirical evidence.

Only proper empirical evidence on why people stick to institutions can decide between these two alternatives (or possibly another third). However, I doubt that the evidence will favour a single account. For I doubt that the persistence of *all* institutions is *either* explained by the presence of an incentive structure *or* by desire-independent reasons for actions.

6. CONCLUSION

I have argued that Hindriks and Guala's functionalist account of the emergence and persistence of institutions suffers from three main problems.

First, their account is unable to make useful predictions on the emergence of specific institutions. Second, their account relies on an inappropriate conception of the benefits of institutions. Because all institutions, whether they exist or could have existed, produce cooperative benefits, I have argued that reference to these cooperative benefits cannot explain why some institutions emerge and persist while others fade out. Third, their account is ahistorical and lacks empirical support. Even if it is adapted to counter the objections highlighted above, for instance by imagining a story of how institutions slowly emerged through a series of self-reinforcing steps, a theoretical account of the emergence and persistence of institutions is not enough and remains hypothetical until it is shown that it can be supported by some historical evidence.

As evolutionary biology has shown, functionalist theories aiming to explain the emergence and persistence of a phenomenon by reference to the ex-post unintentional benefits of this phenomenon, can constitute fruitful, rigorous and solid explanations of natural and social phenomena. However, this article showed that Guala and Hindriks' account of the emergence and persistence of institutions is far from satisfying the conceptual and empirical standards by which it might be judged convincing and able to inform the debates on these issues.

ACKNOWLEDGMENTS

The author wants to thank Richard Endörfer, Christian Munthe, Georg Schmerzeck, Frans Svensson, as well as two anonymous reviewers and the editor of this journal for their insightful comments and suggestions on the article. Research on this article has been made possible by a Marie Curie fellowship (*Project n° 101149764*).

REFERENCES

Aydinonat, N. E., and P. Ylikoski. 2018. "Three Conceptions of a Theory of Institutions." *Philosophy of the Social Sciences* 48 (6): 550–568, URL https://doi.org/10.1177/0048393118798619.

Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press, URL https://doi.org/10.1017/CBO9780511616037.

Binmore, K. 2010. "Game Theory and Institutions." *Symposium: The Dynamics of Institutions* 38: 245–252, URL https://doi.org/10.1016/j.jce.2010.07.003.

Braudel, F. 1988. *La Dynamique du Capitalisme*. Paris: Flammarion, 2014.

Butchard, W., and R. D'Amico. 2015. "Alone Together: Why 'Incentivization' Fails as an Account of Institutional Facts." *Philosophy of the Social Sciences* 45 (3): 315–330, URL https://doi.org/10.1177/0048393115581457.

Elster, J. 1983. *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.

Elster, J. 2015. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences. Second Edition*. Cambridge: Cambridge University Press, URL https://doi.org/10.1017/CBO9781107763111.

Eriksson, L. 2019. "Rational reconstructions and the question of function." *Rationality and Society* 31 (4): 409–431, URL https://doi.org/10.1177/1043463119883959.

Graeber, D. 2011. *Debt the First 5,000 Years*. Brooklyn, N.Y.: Melville House, URL https://doi.org/10.1017/S0010417512000102.

Greif, A. 2006. *Institutions and the Path to the Modern Economy. Lessons from Medieval Trade*. Cambridge University Press, URL https://doi.org/10.1017/CBO9780511791307.

Greif, A., and C. Kingston. 2011. "Institutions: Rules or Equilibria?" *Political Economy of Institutions, Democracy and Voting*, edited by N. Schofield, and G. Caballero, Berlin: Springer, 13–43, URL https://doi.org/10.1007/978-3-642-19519-8_2.

Guala, F. 2016. *Understanding Institutions*. Princeton: Princeton University Press, URL https://doi.org/10.2307/j.ctv7h0sjc.

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

Hédoin, C. 2017. "Institutions, Rule-Following and Game Theory." *Economics & Philosophy* 33 (1): 43–72, URL https://doi.org/10.1017/

S0266267116000043.

Hindriks, F., and F. Guala. 2015. "Institutions, Rules, and Equilibria: A Unified Theory." *Journal of Institutional Economics* 11 (3): 459–80, URL https://doi.org/10.1017/S1744137414000496%20.

Hindriks, F., and F. Guala. 2021. "The Functions of Institutions: Etiology and Teleology." *Synthese* 198: 2027–2043, URL https://doi.org/10.1007/s11229-019-02188-8.

Kitcher, P. 2009. *Living with Darwin: Evolution, Design, and the Future of Faith*. New York: Oxford University Press, URL https://doi.org/10.1093/oso/9780195314441.001.0001.

Larue, L. 2024. "John Searle's Ontology of Money and Its Critics." *The Palgrave Handbook of Philosophy and Money. Volume 2: Modern Thought*, edited by J. Tinguely, Palgrave-MacMillan, 721–741, URL https://doi.org/10.1007/978-3-031-54140-7_36.

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

McManus, C. 2002. *Right Hand, Left Hand: The Origins of Asymmetry in Brains, Bodies, Atoms, and Cultures*. Harvard University Press.

North, D. C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press, URL https://doi.org/10.1017/CBO9780511808678.

Poehler, E. 2017. *The Traffic Systems of Pompeii*. New York: Oxford University Press, URL https://doi.org/10.1093/oso/9780190614676.001.0001.

Searle, J. R. 1995. *The Construction of Social Reality*. New York: Simon and Schuster.

Searle, J. R. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press, URL https://doi.org/10.1093/acprof:osobl/9780195396171.001.0001.

Searle, J. R. 2015. "Status Functions and Institutional Facts: Reply to Hindriks and Guala." *Journal of Institutional Economics* 11 (3): 507–514, URL https://doi.org/10.1017/S1744137414000629%20%20.

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

Smit, J. P., F. Buekens, and S. Du Plessis. 2016. "Cigarettes, dollars and bitcoins – an essay on the ontology of money." *Journal of Institutional Economics* 12 (2): 327–347, URL https://doi.org/10.1017/S1744137415000405%20.

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

Tuomela, R. 2002. *The Philosophy of Social Practices. A Collective Acceptance View*. Cambridge University Press, URL https://doi.org/10.1017/CBO9780511487446.

Tuomela, R. 2007. *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press, URL https://doi.org/10.1093/acprof:oso/9780195313390.001.0001.

Weber, M. 1904. *The Protestant Ethic and the Spirit of Capitalism*. London: Routledge, Translated by Talcott Parsons. With an introduction by Anthony Giddens. 2001.

Wright, L. 1973. "Functions." *The Philosophical Review* 82 (2): 139–168, URL https://doi.org/10.2307/2183766.