Rules, Incentivization and the Ontology of Human Society

Guzmán, Gabriel and Frasser, Cristian

Universidad del Tolima, Universidad del Valle

31 December 2017

Rules, Incentivization and the Ontology of Human Society

Gabriel Guzmán[1] **and Cristian Frasser**[2]

Published in *Philosophy of the Social Sciences*, *Volume: 47 issue: 6, page(s): 440-462*

Abstract

Contemporary discussion about the ontology of society identifies two groups of perspectives. One of them, associated with Searle, includes rules in the inventory of elements that constitute social reality. The other one, associated with Smit, Buekens, and du Plessis, claims that rules can be reduced to more fundamental units. Despite the fact that both perspectives seem equally efficient in describing institutional phenomena, we identify both flaws in the viewpoint that dismisses rules and reasons to prefer the alternative position.

Keywords

Social ontology, Institutions, Rules, Incentivization, Searle

In this paper, we intend to organize the part of the discussion that concerns the elements that must be included in a theory about the ontology of society. More specifically, we aim to say something about the discussion of the place of rules in that theory. It is possible to identify two large groups of perspectives on the role of rules in the ontology of society. One sees rules as the result of the actions of human agents. In this case, as stated by Smit, Buekens, and du Plessis (2011, 2014), institutions (and rules more generally) do not have their own ontology and can be fully understood in terms of agents interacting in a determined environment, guided by incentives in accordance with the formula "S is incentivized to act in manner Z toward X." The other position regards rules as determinants of individual actions and therefore claims that an adequate theory of social ontology must include those rules as fundamental, non-reducible, elements of social reality. In this case, to account for institutions, Searle (1995, 2010) proposes the constitutive rules' formula of "X counts as Y in context C."

It is apropos to consider that Smit, Buekens, and du Plessis and Searle seem equally capable of describing the institutional phenomena covered in their works using their respective formulas. In section 1, we synthesize the manner in which they provide those descriptions. We believe that this occurs simply because it is possible to conceive of an indeterminate number of formulas that can describe such phenomena. In this paper, we do not argue in favor of any kind of formula, nor do we propose a particular alternative. Instead, we note some flaws in the perspective that dismisses rules, as embodied in the work of Smit, Buekens, and du Plessis (2011,

[1] Universidad del Tolima, Ibagué, Colombia
[2] Universidad del Valle, Cali, Colombia.

2014). Our aim is to provide some reasons to prefer the perspective that includes rules in the inventory of elements that constitute social reality over the perspective that claims they can be reduced to more fundamental units.

To strengthen our argument, we use a similar strategy to that employed by Smit, Buekens, and du Plessis (2011, 2014), who combine (i) the use of thought experiments, especially Robinson Crusoe (RC) reasoning, as an appropriate instrument to create imaginary scenarios in which some institutions can emerge without the previous existence of another institutional reality and (ii) the reference to perspectives in game theory and economics to demonstrate the compatibility of their theory and those disciplines. Although we employ this strategy to dispute some of their ideas, our work does not necessarily imply a criticism of the entirety of Smit, Buekens, and du Plessis's proposal. Simultaneously, and even though some of our conclusions bring us closer to the vision according to which rules must be included in an ontological explanation of society, our position cannot be seen as an unconditional defense of the Searlian perspective.[3]

The main arguments in support of our position are organized in the following manner. In section 2, we review the use of RC cases proposed by Smit, Buekens, and du Plessis (2011, 2014) and confront their results with the idea that it would be convenient to see human behavior as rule following. This is advantageous to the extent that it provides us with a way of differentiating human behavior from the kind of biologically determined behavior exhibited by other animal species and that is mostly studied in the natural sciences. In section 3, we continue to analyze RC cases, this time in relation to the idea that to have institutions, there must be language. For us, language can be seen as systems of rules that have the particularity of being necessary to conceive, create, and maintain other systems of rules that then constitute the institutions that characterize human society. In section 4, we attempt to show the compatibility between our idea that rules are necessary in a theory about the ontology of society and new developments in game theory. We note that contrary to the tradition in game theory that seems compatible with Smit, Buekens, and du Plessis, new advances in this discipline suggest that rules are not to be regarded as merely a product of individual actions. Instead, they have a special role to play both in expectation formation and shaping individual behavior. In this section, we also build a counter-example to emphasize the limits of thinking about money in an RC example *à la* Smit, Buekens and du Plessis. Finally, we advance our conclusions in section 5.

1. Two Perspectives on the Role of Rules in Social Ontology

For our purposes, we find it unnecessary to distinguish between rules, norms, conventions and similar concepts. Simultaneously, we regard institutions as systems of rules that individuals follow to coordinate their behavior. We also estimate that from a general point of view, there are two types of explanations about the place occupied by rules in the ontology of society. For one type of explanation, rules (and therefore institutions) are the result of the actions of incentivized agents. In this case, institutional facts can be reduced to physical objects, incentives and actions. The other position is the one for which rules are determinants of individual actions and are ontologically non-reducible. That position defends the notion that a theory of the

[3] We also do not have any commitment to the Searlian distinction between regulative and constitutive rules. For an informative discussion of this issue, see Hindriks (2009) and Hindriks and Guala (2015).

ontology of society must include rules as constitutive parts of that society. In this section, we present the general features of both positions. In this, we summarize part of the theories developed by Smit, Buekens, and du Plessis (2011, 2014) and Searle (1995, 2010). We believe that their works are representative of the aforementioned approaches. In particular, we find that both theories seem equally capable of describing institutional phenomena using the two "formulas" proposed by each of them. On the one hand, there is the structure "X counts as Y in context C," and, on the other hand, there is the alternative "S is incentivized to act in manner Z toward X." For expository reasons, we begin by illustrating the Searlian account.

1.1 "X Counts as Y in Context C"

The Searlian explanation of social reality commences with a distinction between "brute" and "institutional" objects. The snow that covers the top of Mt Everest and a hydrogen atom are instances of brute objects, the existence of which is independent of the presence of human beings in the world. In contrast, a dollar bill and a property contract are examples of institutional objects. Such objects are named "institutional" because the presence of human institutions, such as money and private property are necessary for their existence.[4] For Searle, although institutional objects are often physicochemically instantiated, they cannot be adequately described by terms entirely borrowed from the languages of physics and chemistry. For instance, a citizen is always a human being. Nonetheless, the idea of being a citizen exceeds the physical and chemical descriptions of the human body. For Searle, institutions have their own ontology and require their own explanation insofar as they imply concepts that cannot be satisfactorily reduced to terms drawn from the natural sciences.

According to Searle, a large number of institutions are created through so-called "Declarations." Constitutive rules of the form "X counts as Y in C" can be thought of as a variety of Declarations. If the right person performs a certain speech act under suitable conditions of time and space, this performance suffices to create a new institutional fact. Thus, if a company manager announces to one of his employees, "You're fired," that person ceases to be part of the company. The manager does not describe a pre-existing fact, he creates a new one. To better elucidate the "X counts as Y in C" structure, Searle invites us to imagine a primitive tribe that constructs a wall around its territory. The wall is sufficiently high and strong to keep out foreigners, i.e., the wall performs its job by virtue of its physical properties. Over time, the wall gradually erodes into a mere line of stones. From that moment, the barrier becomes symbolic and is recognized as a territorial border by members of the tribe and foreigners. It is possible to account for the new fact using the "X counts as Y in C" formula: the mark of the stones on the ground (term X) counts as a border (term Y) within the tribal context (C). As noted, it is necessary for term Y of the formula to name something more than the physical qualities of the object denoted by the term X. For instance, regarding a ten-euro bill in my pocket, it is possible to state that certain pieces of paper (X) count as money (Y) in context C. However, satisfying the X term, i.e., being a piece of paper with particular characteristics, is not enough to count as money (Y). In addition, the term X does not specify the causal features capable of making an object act as money, at least not without a collective agreement and the human capacity to impose (status) functions on those objects. Thus, Searle concludes that "[a]n institution is any system of constitutive rules of the form X counts as Y in C" (Searle 2005, 10).

[4] It is worth noting that Searle has been reluctant to talk of institutional objects instead of institutional facts.

1.2 "S Is Incentivized to Act in Manner Z toward X"

- Traffic Lights

One way to understand the proposal of Smit, Buekens, and du Plessis is to analyze their 2011 explanation of traffic lights. From a Searlian perspective, objects that intermittently emit lights in three colors require a collective status function assignation such that they count as traffic lights in a context in which each color counts as a signal that allows the performance of a certain action. Instead, Smit, Buekens, and du Plessis (2011, 5) argue that the example can be fully analyzed in terms of actions and incentives. To this end, they propose replacing the formula "X counts as Y in C" with the heuristic "S is incentivized to act in manner Z toward X." Thus, someone who knows that there are rules and laws that provide pecuniary and criminal penalties for those who violate the rules related to the normal functioning of traffic lights is faced with a set of incentives that make him stop his vehicle when the light is red and continue driving when the light turns green. Therefore, the authors conclude that a distinctive feature of institutional facts is their possession of an associated set of actions that involves a system of incentives. In this way, they take issue with the idea that "institutional facts cannot be reduced to natural facts" (Smit, Bueken and du Plessis 2011, 3).

- Money

In the case of money, Smit, Buekens, and du Plessis (2011, 11) consider it possible to demonstrate that in the same way in which a traffic light is something by which I am incentivized to stop, money is only something that I am incentivized to use to engage in trade. In an RC example proposed by Smit, Buekens, and du Plessis, three individuals (Alex, Bob, and Carol) inhabit an island on which barter is practiced. On the island, there are cigarettes, although no one smokes. One day, Alex publicly announces that henceforth he will accept cigarettes in exchange for the shirts he sells. To impart credibility to his announcement, Alex determines to invest resources in making what would be "a fine cigarette case." Meanwhile, Bob and Carol, who are eager to trade with Alex, have incentives to accept cigarettes as money, which reinforces Alex's commitment. In the end, the incentives become stable, and all of the individuals are willing to use the new money.

- Borders

Regarding borders, Smit, Buekens, and du Plessis (2011, 8) offer an alternative to Searle's wall example and propose another RC scenario. In this case, two men are stranded on an island. One of them (Alex) threatens the other (Bob) with attack if he crosses an imaginary line that divides the island in two. Bob replies in a similar manner. Assuming sufficiently credible intimidation, it is expected that the two subjects have incentives to act in a certain way, i.e., to remain on their side of the island. Thus, a border is created according to the formula "S is incentivized to act in manner Z toward X." In this case, the border is defined as a line that an individual has no incentive to cross. After Alex's threat to Bob, they

> can understand the situation fully by using concepts like 'line', 'crossing', 'probability of getting beaten up', etc. The expressed intentions of the actors, if credible, incentivize the two actors not to cross the line, and a border is created. We need nothing beyond an understanding of the incentives and beliefs of the

two actors in order to grasp the situation fully (Smit, Buekens, and du Plessis 2011, 9).

Smit, Buekens, and du Plessis (2011, 7) underline that in the cases examined, the previous existence of institutions is not necessary for the creation of both incentives and new institutions. Instead, for them, the emergence of an institution is the product of a long action-and-incentive process. Accordingly, institutions are viewed as reducible to terms corresponding to actions and incentives. With regard to the latter, we note that for Searle institutions are systems of rules, and so to our mind, what Smit, Buekens, and du Plessis (2011, 2014) are claiming amounts to saying that rules are ontologically reducible entities that, instead of shaping some human behavior, are the final result of a set of individual actions and incentives.[5]

This overview of the approaches of Searle and Smit, Buekens, and du Plessis is not intended to be exhaustive. So far, we merely wish to remark that in principle, both theories seem to be equally able to describe the kind of phenomena that they analyze by employing their proposed formulas. In the following paragraphs, far from saying that the Searlian approach is entirely right, we attempt to provide numerous arguments that lead us to prefer ontological theories that include rules as one of the constitutive parts of social reality. From this perspective, an adequate inventory of that which constitutes our social reality must include human beings interacting, their environment, and rules that they create to shape and coordinate their behavior and the results of their interactions.

2. RC, Conditioned Behavior and Rule Following

In our view, the idea that rules are constitutive parts of social reality can be related to the idea that for human behavior to be comprehended, it should be regarded as rule following. In this sense, to understand human behavior, it is necessary to identify the rules that come into play in a specific social context of interest. As we will make clear in this section, seeing human behavior as rule following provides us an advantage over the perspective embodied in the work of Smit, Buekens, and du Plessis (2011, 2014). This advantage consists in the fact that it is useful to avoid some confusion. Our perspective is related to a certain way of reading Wittgenstein's ideas. According to this view, human behavior is rule-governed behavior, i.e., we act according to rules, and this marks a crucial difference between the phenomena that occur in human societies and those that belong predominantly to the subject matter of natural science. Specifically, we avoid confusing animal behavior that is biologically conditioned with human behavior that is socially constructed.

In 1758, an object was observed orbiting our sun in an elliptical and retrograde manner. In 1705, Edmond Halley predicted this cosmic event using Newton's Laws. It is expected that the next perihelion of this comet will occur on July 28, 2061. Unlike in this case, we cannot make similar laws or accurate predictions of most of the human phenomena that interest us. In particular, we agree with the idea that human behavior implies movement that can be interpreted and normatively judged. Obviously, according to this view, the motion of the comet is not behavior. From a normative perspective, the idea of following a rule is logically inseparable from the idea of making a mistake. In determining whether someone is following a rule, what the person does should be ponderable. In other words, one should be able to

[5] This point has already been remarked upon by Hédoin (2015).

establish whether what the person does is right or wrong from the perspective of the rule in question. We must consider not only the actions of the person whose behavior is in question (i.e., the candidate for the category of rule-follower) but also the reactions of others to what he does. Thus, "[…] it is only in a situation in which it makes sense to suppose that somebody else could in principle discover the rule which I am following that I can intelligibly be said to follow a rule at all" (Winch 1958, 30). Therefore, the use of the rule only fits within a social context in which there is an external check that can indicate whether my actions truly obey a particular rule.

Incidentally, it is common to find literature that understands animal behavior along the following lines:

> An organism does not need a brain to employ a strategy. Bacteria, for example, have a basic capacity to play games in that (1) bacteria are highly responsive to selected aspects of their environment, especially their chemical environment; (2) this implies that they can respond differentially to what other organisms around them are doing[.] (Axelrod and Hamilton 1981, 1392).

We believe that this way of speaking is compatible with the type of analysis typically involved in the use of thought experiments in the vein of Smit, Buekens, and du Plessis (2011, 2014). We could theorize about the world of bacteria by saying that they interact with one another and with their environment following incentives in a similar manner to how fictional humans interact to create institutions guided by incentives in RC examples. Whereas it is clear that Halley's Comet does not follow a strategy, it seems fairly reasonable to assume that bacteria, which are incentivized to obtain food and avoid being eaten, create strategies (such as hiding to ambush their prey or avoiding being caught by predators) in a similar way to that of the characters in RC examples, who react to incentives from their environment to form institutions in a scenario while lacking any previous institution. Nevertheless, we see that this way of proceeding can be misleading. Thus, for example, if our aim is to understand and not just to describe the behavior of children playing hide-and-seek, it is necessary to know the rules involved in that game. In contrast, to understand the world of bacteria does not require a similar understanding because bacteria do not follow rules. The behavior of bacteria cannot be analyzed using this scheme. These organisms do not act according to socially established rules for the simple reason that they lack a society. Bacteria act according to naturally determined mechanisms. In this regard, the behavior of bacteria is closer to that of Halley's Comet than to that of children playing hide-and-seek.

In response to Axelrod and Hamilton (1981), Ramos (2011) states that there are clear cases in the animal biological world in which there is no rule following, only behavior that is causally explainable. From this perspective, except metaphorically, bacteria do not behave according to strategic rules, they do not play games, and insects such as bees have no real queens that can be dethroned. Instead of using the expression "rule-governed behavior," in such cases, it is better to refer to biologically conditioned behavior. Additionally, there are clear examples of highly rule-governed societies, such as republics. Nevertheless, numerous gray cases remain (for example, very primitive humans or other higher mammals) in which it becomes difficult to separate instinct from will and the normative from the causal (Ramos 2011). For us, theoretical proposals such as that of Smit, Buekens and du Plessis (2011, 2014) can confuse conditioned behavior (such as that of bacteria), phenomena comparable to gray cases, and the particularities of evolved, highly rule-governed human societies.

To understand a little more about those difficult gray cases, we can imagine a genealogy of contemporary borders based on the barriers constructed by prehistoric humans to repel attacks. Such barriers can only fulfill their function to the extent that their physical properties allow it. In the animal world, we can state that foreign packs are incentivized not to cross physical marks, given incentives based on the local pack's ability to cause physical harm to invaders. Whereas the dividing line is not clear, unlike our borders, the primitive barrier and the marks of the wolves, rather than being institutions in and of themselves, appear to be proto-institutions. As proto-institutions, they can be viewed on the one hand as physical objects capable of preventing or favoring the occurrence of some physical facts and on the other hand as capable of evolving into institutions in which the physical element generally becomes increasingly less relevant.

Properly examined, it appears that when an RC example such as that mentioned in section 1.2 is proposed to account for the emergence of a border (or any institution) without the mediation of previous institutions, the most likely result is a situation in which what is similar in the first instance in its nature to our human institutions ends up resembling more simple brute facts (as is the case for most animals) or a proto-institution (in the case of certain more evolved species that have what could be called proto-societies and proto-languages). For us, what is described in the island scenarios of Smit, Buekens and du Plessis (2011, 2014) do not represent the emergence of institutional reality but, generally, a set of brute facts and proto-institutions that could possess some elements in common with human institutions.

Taking liberties with language, we can speak of the border of the territory inhabited by a bacterium. In this case, however, everything is supported by the bacterium's physical ability to defend its space. Nevertheless, although the use of contemporary borders often involves the intervention of physical phenomena (such as the territory over which the imaginary line runs and the barbed wire that protects certain routes), the most important characteristic (at least the characteristic that distinguishes our species) is the existence of a wide range of another kind of phenomenon that transcends the physical attributes of the objects that support them and give shape to a complex web of institutional objects and facts such as states, immigration laws, passports, customs, international trade, invasions, and the like. To a certain extent, what occurs in a territory inhabited by packs of wolves can be compared with what occurs on the border that separates Mexico from the United States. That is, we can use the word "border" when accounting for how the territory is divided among the wolves, and we can also use the word "border" to talk about an imaginary line, geographical features, or barriers placed by humans to separate the territories of those two countries or talk about drug trafficking, illegal migration, cultural expressions, political tensions, etc. Nevertheless, it is certainly true that the same sets of phenomena do not have the same meaning when the same word in English, "border," is used to describe what occurs in each situation.

We have insisted that the two formulas seem equally efficient to describe the nature of society. However, to understand human behavior, it appears convenient to see it as rule following. This has the advantage of making it possible for us to judge if the movement of someone crossing an imaginary line corresponds to the occurrence of an invasion, a legal migration, or merely a simple accident. Let us think, for instance, about someone who ignores a no-trespassing sign. For us to judge what could have occurred, we must know the appropriate context. This allows us to distinguish between a situation in which someone crossed the imaginary line because she was blind, because she could not read, because she was distracted,

because it was too dark, or simply because she was a thief. In cases such as these, there is considerable room for interpretation. In gray cases, such as those of wolf packs sharing territories, it is more difficult to specify the scope of a rule-following analysis. Sometimes, it seems possible to observe in the behavior of wolves similarities to the behavior of human beings, for instance, when they decide to invade others' territories because of some intention. Nonetheless, there are also many situations in which the adequate explanations of those animals' behavior are strictly biological. For now, we cannot say much more about the characteristics of these gray cases. In sharp contrast, we can describe the behavior of a bacterium in relation to its environment and say, as *une façon de parler*, that under certain circumstances it crosses a border. We can see the bacterium moving and crossing the imaginary line and explain what it is doing by saying that it is being incentivized for some reason. Nevertheless, it does not make much sense to say something like the bacterium invades a country, or it is simply a tourist in a foreign country. In this case, we cannot state that the bacterium is following any rule for its behavior because it is fully biologically conditioned and therefore, the appropriate explanation must be found in the natural sciences. We think theoretical work in the social sciences should take note of this to avoid the kind of confusion described above.

3. Rules, Language and Institutions. What Else Do RC Examples Tell Us?

Part of what has been previously said about rule following can be extended to the case of language. For Searle (1969), speaking a language is to participate in a form of highly complex, rule-governed behavior. "To learn and master a language is (inter alia) to learn and to have mastered these rules" (Searle 1969, 12). This approach is compatible with some scientific theories. In particular, in linguistics, we highlight the perspective developed by Port (2010), who defends the idea that *Homo sapiens* somehow managed to develop complex cultures capable of creating a wide range of technologies suited to the environment. These technologies included, e.g., grammar, religious practices, education and government, which, combined with other technologies related to productive activities necessary for survival and reproduction, such as food production, together with everything that this implies regarding the development of tools, accumulated knowledge, skills and social conventions, have shaped the culture of human societies. In contrast to the idea that language is a form of knowledge, for this author, "[…] a language is a kind of social institution, that is, a partially structured system of conventions created by a community of speakers and refined over generations. It is a technology developed by a community for coordination of behavior" (Port 2010, 313).

Port (2010, 319) offers an explanation that is intended to be compatible with findings in fields such as anthropology and evolutionary biology. In this regard, he hypothesizes that the first known cultural flowering, which occurred one hundred thousand years ago, must have coincided with the emergence of language. Subsequently, permanent cultural change occurred, as demonstrated by the continuous adaptation and specialization of tools and practices to suit the specific conditions of each environment. According to this view, society appears as a complex adaptive system capable of creating structures, such as lexicons and phonologies, that are in and of themselves conventions created by a group over time.[6] In this sense, agents appear in the system as imitators and users of speech conventions established by the community.

[6] In this case, as was pointed out at the beginning of this paper, it is unimportant for our purposes to distinguish between conventions and rules.

In what follows, we note more difficulties and implications related to the conception of RC cases used to imagine scenarios in which an institution emerges in a context free of any other previous institutional reality. Specifically, assuming language is an institutional phenomenon, in the following subsection, we attempt to demonstrate that the evolution of institutions seems particularly improbable without the existence of language. If this is correct, perhaps social science should pay more attention to the ruled and linguistic character of human institutions to understand their nature.

3.1 RC Examples without Language

When we look at the RC examples provided by Smit, Buekens, and du Plessis, we cannot ignore the permanent presence of language. In addition to the cases mentioned in the first section, this is clear in every segment of their work that employs this explanatory device.[7] For instance, they write the following:

> Consider ten people who independently and simultaneously wash up on a desert island and who all speak a version of English that contains no institutional fact terms. Due to the practical gains to be had from sticking together and their belief in the edifying effects of human society they live in the same place, talk, interact, and so forth. (None of this necessarily presupposes the existence of an institutional fact, as Searle uses the term 'institution'.) (Smit, Buekens, and du Plessis 2014, 1823).

Smit, Buekens, and du Plessis do not intend to exclude language from their explanation. Nevertheless, we believe that if they did attempt to do so, their effort would be unsuccessful. This helps us evince the role of language as systems of rules and a necessary condition for the existence of institutions. Returning to the border example, the situation described by Smit, Buekens, and du Plessis (2011, 9) cited in section 1.2 is viable insofar as there is a shared language of which those concepts are a part. Before being stranded on the island, Alex and Bob lived in a society in which they learned a language with words such as "line," "crossing," "probability," "border" and "threat." In this case, a kind of "theory-ladenness" enables someone to view a line as a border. A dog may see the same line but has no "theory-ladenness" that enables it to "see it as" a border. This feat is something that a dog cannot perform. A dog cannot recognize or imagine the line that demarcates a border simply because a dog does not have a language and therefore lacks the necessary concepts.

Once again, let us consider the case of the emergence of a border on the island of Alex and Bob. It is possible to conceive that Alex and Bob are abandoned on the island before learning a language, and they survive and behave in such a way that (as with certain species) the territory is delimited using scent-marking and physical aggressions until the island is divided in two. An outside observer could verify that there is an imaginary line that divides the island in two and that Alex only attacks Bob when the latter invades Alex's half of the island. Thus, from the perspective of Smit, Buekens, and du Plessis (2011, 2014), an institution has been formed that is entirely supported by the actions of the agents, along with some incentives.

[7] This is the case in Smit, Buekens, and du Plessis (2011, 8, 16 and 18). Likewise, the presence of language is important in other types of mental exercises, for instance in Smit, Buekens, and du Plessis (2014, 1822).

In principle, imagining this situation in the absence of language does not seem problematic and apparently, the imaginary line can be understood as a border. However, as we stated above, in this scenario, we can confuse different kinds of phenomena. Instead of highlighting the most fundamental elements of a society, this thought experiment recalls situations that are more similar to animal behavior cases, highly conditioned by instinct, or perhaps to proto-institutions exhibited by animal species that possess what could be considered proto-languages.

The border example has been constantly emphasized because it provides us with a degree of progress in the exercise of imagining an RC scenario lacking in prior institutions, including the institution of language. Nevertheless, when one considers something like money, this project does not seem to leave any additional room for maneuvering. First, the extreme case of a single agent conceiving money seems absurd, as Wittgenstein (1956, §45) suggests concerning the idea of an isolated individual engaging in commercial practices. However, the situation of two agents without language inhabiting an island and creating money does not seem plausible either. As we have insisted, certain institutions have a remote past that makes them comparable to observed phenomena among other types of societies, for instance, the similarity between a medieval king and the alpha male of a wolf pack. For their part, other institutions seem to be different and to have no known parallel outside the human world, which seems to be the case with money. Part of the explanation would be that to have money, it is necessary to have language. Otherwise, it is hard to conceive that a piece of paper, with low or no "use-value", could be understood as exchangeable for a vital good. We cannot offer a thought experiment in which some creatures lacking any kind of language interact until the point of creating an institution such as money. It is not that in the scenario created by Smit, Buekens, and du Plessis, we cannot imagine Alex obliging Bob to accept cigarettes as money by making use of force and without using a single word; it is that without a language, it seems implausible to conceive of the very concept of money. Simply without language, Alex cannot see beyond the cylinder of paper stuffed full with tobacco.[8]

In accordance with the above, the standard position in cognitive science postulates that concepts are mental representations (Thagard 2005, 59). This conception was originally proposed by the Representational Theory of the Mind (RTM), according to which thinking occurs in an internal system of representation. A version of RTM considers concepts not only as mental representations but also as the constituents of propositional attitudes such as beliefs and desires (Margolis and Laurence 2007, 563). It is this version of RTM that is now widely embraced in both cognitive science and the philosophy of mind. One of the advantages of this account is its ability to explain the productivity of thought; that is, leaving aside certain limitations of memory and attention, humans seem to have a formidable capability to entertain an infinite number of thoughts. Humans are indeed capable of understanding sentences that they have neither heard nor uttered before but which are formulated under the rules of their own language.

The relation between concepts, thought, and language remains a subject of intense debate among philosophers and scientists. However, numerous philosophers

[8] It appears feasible to teach capuchin monkeys to use money, although only in its role as a medium of exchange (Chen, Venkat and Santos 2006, 517). Without wading into the controversy over the differences among concepts in humans and mere representations in animals, we think this can occur because humans can transmit some practices to animals with a certain degree of intelligence, in particular those that are highly social and have proto-language.

and scientists agree that some kind of thought can occur prior to language (Bermúdez 2003, 3; Hurlburt and Akhter 2008, 1364; Searle 2010, 61). This seems to fit with the evidence of some animals capable of having mere representations (Clayton, Bussey and Dickinson 2003, 685; Emery, Dally and Clayton 2004, 37). Although it appears possible in principle to have some representations and thoughts without language, the types of representations and thoughts entertained by non-linguistic creatures seem to be restricted to a limited range. Bermúdez (2003, 150) notes that, for instance, non-linguistic creatures are not capable of thinking about thoughts. We are not interested in entering into the philosophical and scientific controversy surrounding this evidence. Instead, our interest is to notice that although it is possible to entertain certain representations (and hence some thoughts) without language, not all kinds of representations are allowed without language. Indeed, when we move to the case of humans, we find evidence that some kinds of human thinking do require a linguistic system (Carruthers 2002, 657; Shusterman and Spelke 2005, 95). Put differently, concepts involved in some human tasks require language-based thinking. Unlike non-linguistic creatures, humans with a finite base of symbols and words can exhibit an unlimited capacity to entertain novel and diverse thoughts. We argue that the kind of concepts involved in creating our institutional world requires thinking that occurs through a vehicle, and that vehicle is language.[9] By employing a linguistic medium, humans can use concepts that are expressed linguistically, enabling them to create institutions such as money.

Studying the RC cases used by Smit, Buekens, and du Plessis, we find a particularity, namely, the fundamental role of language in the emergence and functioning of institutions. We think that this can be seen as an argument in favor of the idea that a complete ontological theory of human social reality, useful for understanding and not merely describing purposes, should include rules. From our point of view, language are systems of rules that are necessary for conceiving and maintaining institutions. For now, it seems promising to recognize the importance to our species of having a language capable of creating a part of reality that is a subject matter of the social sciences. To what extent, for instance, thinking about the nature of market liquidity (a complex network of debts, meaning promises to pay) in these terms helps us better understand that phenomena such as financial crises and the design of monetary institutions remain to be explored.

4. Rules, Game Theory and Money

So far, we have made two arguments in favor of the idea that institutional reality has agents interacting in a certain environment guided by incentives, physical objects, and various rules as its constitutive elements. On the one hand, we can say that language and all human behavior in general are rule-governed behavior. Despite the existence of gray cases, this marks a notable difference from other species' behavior. On the other hand, we showed that there are good indications that institutions can only flourish in a scenario in which there must be language, the latter being systems of rules. In this section, we seek to provide a third argument. We appeal to game theory to illustrate that rules can change the nature of the game played by individuals. Accordingly, rules can produce individual actions and

[9] Recently, Hindriks and Guala (2015), who have been critical of the Searlian account, seem to concede this point: "Non-human animals solve coordination games using correlation devices, but animals do not have institutions." Later, they state that "[i]n the case of humans, we say that they can follow different rules. These rules are representations in symbolic form of the strategies that ought to be followed in a given game" (Hindriks and Guala 2015, 465).

outcomes that would be unintelligible in the absence of such rules. Instead of being the end of a long action-and-incentive process, rules make the existence of that process possible. Leaving aside for now our argument against the possibility of having institutions without language, we propose a counter-example to illustrate the limits of thinking about money only as a product of individuals, incentivized actions, and physical objects.

Smit, Buekens, and du Plessis claim that one virtue of their approach is its compatibility with game theory and standard economic theory. The reason for this confidence seems to be that their ontological explanation based on actions and incentives is in line with the ontological commitments of standard game theory. Without further details about the state of standard economic theory and game theory, Smit, Buekens, and du Plessis write,

> The incentivized action view allows for easy vertical integration with other theories. Our theory of promising as self-incentivization, for instance, fits seamlessly into game theory and mainstream economic theory in general. Game theory is antecedently committed to the existence of agents with various beliefs, courses of action and incentives to perform such actions. The incentivized action view attempts to show that all talk of money, borders, promises and the like can be accommodated into game theory as simply being yet more talk about actions and incentives (2014, 17).

We believe that such trust is not well founded, not because game theory conflicts with individual actions and incentives but because it neglects theoretical difficulties confronted by game theory that are indeed linked in some way with those game theory elements that Smit, Buekens, and du Plessis judge as compatible with their own ontological account. Thus, for example, the standard version of game theory has had problems in becoming a theory capable of successfully explaining and predicting important laboratory results and real-world interactions. Furthermore, game theorists often encounter the so-called problem of indeterminacy. That is, many games have more than one equilibrium and therefore, it is not possible to identify a unique outcome, which amounts to saying that rational agents acting according to their incentives cannot pick a strategy in a real-world context.

To confront such difficulties, some game theorists have opted to include social objects (institutions, norms, conventions or simply rules) to end indeterminacy by choosing one among multiple equilibria. Likewise, once such objects are introduced into the analysis, some of the results achieved appear to fit well with the data collected in controlled environments. Hédoin (2015) shows two alternative approaches that call into question the alleged compatibility between game theory and the incentivized action view. The first approach is derived from psychological and behavioral game theory, which sees rules as normative expectations; the other approach, epistemic game theory, considers rules to be correlating devices. In this section, we only refer to the results of the first approach.

According to Hédoin (2015), psychological and behavioral game theory, unlike the standard account that seeks to provide normative or prescriptive guidelines about what rational agents ought to do, elaborates methodological and theoretical proposals that tend to reflect what people actually do in strategic interactions. Within this approach, Hédoin (2015) highlights the works of Rabin (1993), Sugden (2000), and Bicchieri (2006). In the first case, Rabin presents a model in which some phenomena can be explained through a preference for fairness and reciprocity. If the

other player is expected to be kind, *ceteris paribus*, I will act in such a way that my actions correspond to that expectation, thereby reciprocating the expected kindness with a behavior considered equally kind by the other. Sugden's proposal, which is focused on resentment aversion, follows a similar line of argument; he argues that conventions gradually acquire a normative force until they become well-established social norms. For Sugden, humans have a psychology according to which we resent those who frustrate our reasonable expectations and avoid actions that generate others' resentment toward us. For its part, in Bicchieri's model, social norms are founded on the conditional preference for conformity. Social norms are closely connected to practices to which we conform because we expect others to conform to them and believe that others expect us to do likewise. For Hédoin (2015), the explanation of rules as normative expectations is characterized by recognizing that as the payments are a function of the players' beliefs, especially those of second order, the nature of the game changes in accordance with the players' beliefs in a given context. Assuming that players' beliefs are formed from social rules, one must conclude that changes in rules change the game. Bicchieri shows formally that social norms can alter the game and that the very nature of the game (whether it is a game of coordination, mixed motives or Bayesian) depends on the social norms that exist and are actually followed by the agents. In this regard, the agents' behavior depends on expectations that would not exist in the absence of the norm. Put differently, individual actions are norm-driven, and therefore, they cannot be conceivable without the existence of such norms.

To take the argument a little bit further, let us employ RC reasoning again. Leaving aside the difficulties mentioned above for the existence of money in a context lacking previous rules, let us imagine that Alex, Bob and Carol use cigarettes as money on the island. We create a scenario where they are incentivized to dismiss the use of money, and then we show how the introduction of a rule reinstalls money as an exchange technology. Alex, Bob and Carol make their production and consumption decisions during the week. On the last day of the week, they must also decide whether they will hold positive money balances for the next week. One day they receive a message delivered by a homing pigeon indicating that in exactly six weeks, a boat will come to rescue them. Assuming they have a calendar system to count the weeks, it is not hard to see that the immediate outcome is a demonetization of the economy generated by a change in individual beliefs caused by this message. Before receiving the message, everybody had incentives to hold money because production needed to be funded at the beginning of the next week. Consequently, the island's economy remains a monetary economy. Once Alex, Bob and Carol learn that they will leave the island at the end of the sixth week, the incentives to hold money do not work anymore. At the beginning of the final week, everybody knows that holding money is not a wise decision because there will be no trade on the island in the following week. Thus, Alex, Bob and Carol will attempt to dispose of all their money during the last week. However, since nobody wants to hold positive money balances, no one has an incentive to accept money in exchange for goods—remember also that in this RC case, by construction, Alex, Bob and Carol do not smoke. In other words, the island becomes a non-monetary economy during the last week. It is now worth examining what occurs during the next-to-last week. In the fifth week, everybody knows that in the sixth week, no one will accept money, so Alex, Bob and Carol have no incentive to hold money at the end of that week either. As a result, the island will also be effectively demonetized in the fifth week. The same process goes

for the other weeks, with a final outcome of a moneyless economy from the moment that Alex, Bob and Carol received the message.[10]

The moral we draw from this simple thought experiment is that actions, incentives, and physical objects might not be enough to account for money. Once Alex, Bob and Carol know that their stay on the island is limited to a few weeks, their beliefs and incentives change, thus demonetizing the economy. To explain the permanence of money until the day they leave the island, a social rule could be considered. Thus, for example, all those who lived on the island must pay a tribute in cigarettes to a central authority before getting on the boat and leaving.[11] Now, Alex, Bob and Carol do have an incentive to hold money until the last week: taxes. Accordingly, the economy does not demonetize because of the tax rule. As in the foregoing cases, the tax rule modifies individual beliefs and gives rise to incentives that lead to novel individual actions. Without the presence of such a rule, Alex, Bob and Carol, guided by their individual incentives, would choose a non-monetary economy. Aside from our claim against the emergence of an institution in the absence of language, in this context, money is a product of a prior social rule (the payment of taxes), actions, incentives, and physical objects.[12]

5. Conclusions

In this paper, we separate visions of the role of rules in an explanation of the ontology of social reality into two groups. From one point of view, rules are the result of incentivized actions; and in its theory of society, rules are not necessary as they can be reduced to an explanation fully based on those incentivized actions. From the other point of view, rules are determinants of human actions, and they are non-reducible entities. Regarding this discussion, we can say that both perspectives seem equally capable of describing social phenomena that employ their formulas. The first perspective, which is associated with Smit, Buekens, and du Plessis (2011, 2014), uses the formula "S is incentivized to act in manner Z toward X." The second perspective, as in Searle (1995, 2010), uses the formula "X counts as Y in C."

Although both approaches seem equally efficient in describing institutional phenomena, when thinking about the labor of understanding rather than describing social nature, it would be helpful to avoid removing rules from a theory of society. For instance, to accept that rules are part of the inventory of the social world and that human behavior is rule-governed behavior helps us escape the confusion between different kinds of phenomena. In particular, it permits us to distinguish between

[10] Technically speaking, in this counter-example, money does not display a positive price. The first economist to note the problem of the positivity of money price was Frank Hahn (1966). Since then, this has been one of the most elusive problems in monetary theory.

[11] This method of confronting the positivity of the price of money has been explored by Starr (1974).

[12] Economists have certainly faced many difficulties when trying to find a rationale for the emergence of money, lacking a prior institutional context. In other words, it is difficult to provide an account of money within the domains of economic analysis. This problem was expressed by the economist James Tobin: "The principal reason, then, that Treasury bills [for example] are not media of exchange is that they are not generally acceptable. This unsatisfactory circular conclusion underlines the essential point, that general acceptability in exchange is one of those phenomena—like language, rules of the road, fashion in dress—where the fact of social consensus is much more important and much more predictable than the content. Whether or not an asset is a generally acceptable means of payment will have to be taken as a datum determined by legal institutions and social conventions, rather than as something to be explained by economic theory" (Starr 1989, 293).

biologically conditioned behavior exhibited by some species lacking society and human behavior, which is socially conditioned and able to be normatively judged.

With the intention of arguing in favor of the inclusion of rules in an explanation of human society and its institutions, we use similar tools to those used by Smit, Buekens, and du Plessis. The result of the analysis of RC cases prompts us to think that it is not possible to conceive of institutions in an environment completely lacking language. Without language, we cannot conceive of, create or maintain institutions. Some gray cases connect humans and their prehistoric ancestors through evolution. In these cases, it is possible to think about proto-societies with proto-languages that can create proto-institutions. This kind of phenomena can also be seen in other highly evolved animal species.

For us, a good part of the strength of the theoretical perspective of Smit, Buekens, and du Plessis lies in its alleged compatibility with standard economics and standard game theory. However, it is well worth recognizing the existence of alternative viewpoints in game theory that offer actual solutions to problems that have not been overcome by traditional approaches, as is the case for the problem of indeterminacy. If we are right, it seems that the study of the possible implications of developing the theoretical role of rules in the functioning of human society deserves more attention. For instance, the theory proposed by Smit, Buekens, and du Plessis is compatible with traditional perspectives in economics such as those endorsed by Adam Smith, Carl Menger and some search models of money. From our point of view, monetary economists have been particularly reluctant to include the institutional context in their explanations, and we consider that a change in this practice could have interesting theoretical implications for monetary economics.

Many current questions related to rules will be answered as a result of advancements in various scientific disciplines. Nevertheless, we want to hypothesize that it is possible that some of the limitations of social sciences are attributable to the lack of progress in understanding the place and nature of rules in the functioning of society. In the case of economics, it could be true that one part of the limitations in the development of satisfactory responses to some theoretical and practical problems is practitioners' relative lack of interest in understanding the role of rules, mainly institutions, in market economies.

References

Axelrod, Robert, and William Hamilton. 1981. "The Evolution of Cooperation." *Science* 211 (4489): 1390-1396.

Bermúdez, José Luis. 2003. *Thinking without Words*. New York: Oxford University Press.

Bicchieri, Cristina. 2006. *The grammar of society: The nature and dynamics of social norms*. Nueva York: Cambridge University Press.

Carruthers, Peter. 2002. "The Cognitive Functions of Language." *Behavioral and Brain Sciences* 25, nº 06: 657-674.

Chen, Keith, Lakshminarayanan Venkat, and Laurie Santos. 2006."How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior." *Journal of Political Economy* 114, nº 3: 517-537.

Clayton, Nicola, Timothy Bussey, and Anthony Dickinson. 2003. "Can animals recall the past and plan for the future?" *Nature Reviews Neuroscience* 4: 685-691.

Emery, Nathan, Joanna Dally, and Nicola Clayton. 2004. "Western scrub-jays (Aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics." *Animal Cognition* 7, nº 1: 37-43.

Hahn, Frank. 1966. "On some problems of proving the existence of an equilibrium in a monetary economy." *The Theory of Interest Rates.* F. Hahn and F. Brechung, eds. Macmillan, London and Basingstoke.

Halley, Edmond.1705. *Synopsis Astronomia Cometicae.* London: John Senex.

Hédoin, Cyril. 2015. "Accounting for constitutive rules in game theory", *Journal of Economic Methodology*, 22, 4: 439-461.

Hindriks, Frank. 2009. "Constitutive Rules, Language, and Ontology." *Erkenntnis* 71, no. 2: 253-275.

Hindriks, Frank, and Francesco Guala. 2015. "Institutions, rules and equilibria: a unified theory." *Journal of Institutional Economics* 11, nº 3: 459-480.

Hurlburt, Russell, and Sarah Akhter. 2008. "Unsymbolized thinking." *Consciousness and Cognition* 17: 1364-1374.

Margolis, Eric, and Stephen Laurence. 2007. "The Ontology of Concepts—Abstract Objects or Mental Representations?" *Nous* 41, nº 4: 561-593.

Port, Robert. 2010. "Language as a Social Institution: Why Phonemes and Words Do Not Live in the Brain." *Ecological Psychology* 22: 304-326.

Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics", *The American Economic Review*, 83: 1281–1302.

Ramos, Jaime. 2011. "Two Notions of Normativity." Lecture, Fondation Hugot-Collège de France, Paris, 6 December, 2011.

Searle, John. 1969 (1999). *Speech Acts: An essay in the Philosophy of language.* New York: Cambridge University Press.

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

Searle, John. 2010. *Making The Social World: The Structure of Human Civilization.* New York: Oxford University Press.

Shusterman, Anna, and Elizabeth Spelke. 2005. "Language and the Development of Spatial Reasoning." *The Innate Mind: Structure and Contents*, Peter Carruthers,

Stephen Laurence and Stephen Stich eds., 89-106. New York: Oxford University Press.

Smit, J.P., Filip Buekens, and Stan du Plessis. 2011 "What Is Money? An Alternative to Searle's Institutional Facts." *Economics and Philosophy* 27: 1-22.

Smit, J.P., Filip Buekens, and Stan du Plessis. 2014. "Developing the Incentivized Action View of Institutional Reality." *Synthese* 191: 1813-1830.

Starr, Ross. 1974. "The price of money in a pure exchange monetary economy with taxation. *Econometrica*, 42, No. 1: 45-54.

Starr, Ross. 1989. *General Equilibrium Models of Monetary Economies. Studies in the Static Foundations of Monetary Theory.* San Diego, CA: Academic Press.

Sugden, Robert. 2000. "The motivating power of expectations", *Rationality, rules, and structure,* Nida-Rümelin J. and W. Spohn eds. 103–125. Norwell, Kluwer Academic Publishers

Thagard, Paul. 2005. *Mind. Introduction to Cognitive Science.* Cambridge, Massachusetts: The MIT Press.

Winch, Peter. (1958) 1960. *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge & Kegan Paul.

Wittgenstein, Ludwig. 1956 (1978). *Remarks on the Foundations of Mathematics.* G. H. von Wright, R. Rhees and G. E. M. Anscombe (eds.), G. E. M Anscombe (trans.), Oxford: Blackwell.