

# Rule-Following and Constitutive Rules: A Reconciliation

Cyril Hédoin\*

*University of Reims Champagne-Ardenne (France)*

Version 2.0: 19<sup>th</sup> March 2017

**Abstract:** This article contrasts two broad approaches of rules and institutions in the philosophy of social sciences: on the one hand, the approach that deals with institutions and institutional facts in terms of rule-following behavior, on the other hand the approach that characterizes institutions in terms of constitutive rules. This latter approach builds on a distinction between two kinds of rules – constitutive and regulative rules – that is alien to the rule-following account. I argue that this distinction is inadequate and that all rules contribute to the construction of the social reality by being followed – and only followed. The main contribution is to suggest a distinction between “constitutive” or “essential” rules in terms of intentional rule-following behaviors. Constitutive rules are rules that must be followed by individuals taking part to an institutional practice because of the value ascribed to the practice by these individuals. As a consequence, whether or not a rule is constitutive is an ontologically subjective (rather than objective) but epistemically objective fact.

**Keywords:** Rule-following – Constitutive Rules – Institutional Facts – Intentional attitudes

## 1. Introduction

The concepts of rules and institutions naturally occupy a central place in social ontology and the philosophy of social sciences. Rules and institutions are distinguishing features of human societies and any account of the nature of social reality has to figure out their role and their functioning. There are however many ways to account for rules and in particular how they are related to the individuals’ behavior in a population.

This article contrasts two broad approaches of rules and institutions in philosophy of social sciences: on the one hand, the approach that deals with institutions and institutional facts in terms of rule-following behavior, on the other hand the approach that characterizes institutions in terms of constitutive rules. This latter approach builds on a distinction between two kinds of rules – constitutive and regulative rules – that is alien to the rule-following account. At the heart of the distinction between constitutive and regulative rules is the idea that constitutive rules are not rules to be merely followed; rather, they create the social reality by ascribing status to

---

\* Full Professor of Economics, *UFR de sciences économiques, sociales et de gestion* and economics and management research center *REGARDS*, University of Reims Champagne-Ardenne.  
Contact: [cyril.hedoin@univ-reims.fr](mailto:cyril.hedoin@univ-reims.fr)

entities or states of affairs. The philosopher John Searle has been the major proponent of the distinction between constitutive and regulative rules and has argued that the former are at the basis of the social reality. The notion of rule-following has been particularly developed in the late writings of Ludwig Wittgenstein and has found many applications in the social sciences, for instance in game-theoretic accounts of institutions in economics [(Hindriks and Guala 2015); (Hédoin 2015)].

I argue in this article that Searle's various criteria for the distinction between constitutive and regulative rules are unsatisfactory because they depend on the description of rules. Depending on one's preferred descriptions, all rules can be seen as constitutive or regulative. In particular, regulative rules are generally stated in terms of behavioral descriptions while constitutive rules are stated in terms of status descriptions. However, both kinds of descriptions are formally equivalent. Actually, all rules are rules to be followed and nothing more. The main contribution of the article is then to suggest a characterization of "constitutive" or "essential" rules in terms of intentional rule-following behaviors. Constitutive rules are rules that must be followed by individuals taking part to an institutional practice because of the value ascribed to the practice by these individuals. As a consequence, whether or not a rule is constitutive is an ontologically subjective (rather than objective) but epistemically objective fact.

The rest of the article is organized as follows. Section 2 briefly defines the key concepts. Sections 3 and 4 discuss Searle's three criteria for distinguishing constitutive and regulative rules and argue for their inadequacy. Sections 5 and 6 build on Wittgenstein's and Rawls' practice conception of rules and argue that there are rules that are constitutive in the sense that individuals cannot take part into a practice without intentionally following them. Section 7 responds to potential objections to my account and concludes.

## **2. Institutions and Constitutive Rules**

Any account of social ontology must deal with a range of concepts such as institutions, institutional facts, institutional objects, social facts, and so on. Definitions vary within the literature but institutional facts are generally regarded as a subcategory of social facts. I will adopt a broad view of a social fact as *any set of repeated interactions between intentional beings that give rise to some regularity at the behavioral level*. This definition encompasses interactions between humans leading to some behavioral pattern (think of the road traffic) but also the complex organizations of collectives of insects or animals.<sup>1</sup> Institutional facts are more specific. They involve the existence of institutions. Examples of institutions are markets, money, marriage, language, corporations, etc. Institutional facts are then social facts that take place within a practice corresponding to an institution. Consider the following facts:

- a) 'John bought a car.'
- b) 'I have a twenty dollars bill in my wallet.'

---

<sup>1</sup> Some authors like Searle (1995, 121) impose a condition of collective intentionality on social facts. Since I do not intend to discuss the issues related to collective intentionality, I omit this condition here.

c) ‘He has hit a home run in his last baseball game.’

These facts presuppose the existence of an institution (respectively, a market, money and baseball): the fact of buying something must take place on a market, the fact of having a twenty dollars bill in my possession necessarily implies the institution of money and hitting a home run is meaningless but in a baseball game, the latter itself presupposing the institution of baseball.<sup>2</sup> Institutional facts are social facts according to the above definition: they depend on the repeated interactions between intentional beings and occur on a regular basis. For instance, the very fact of having a *twenty dollars bill* in your wallet has meaning only through the way this piece of paper allows you to interact with others on a repeated and predictable basis. Institutional facts add something else however. Compare the collective behavior of a stack of fishes with the behavioral pattern that characterizes a baseball game. In both cases, you will be able to identify a behavioral pattern such that it is in principle possible to predict that in some specific circumstances, the whole collective will behave in some specific way. What distinguishes the baseball game however is that the behavior of each player (and thus the collective behavioral pattern) results from the fact that the players are *intentionally following some rule*. For the moment, this may seem obscure because the very notion of “intentionally following a rule” may be interpreted in quite different ways. This will be clarified below.

If institutional facts consist in rule-following behaviors, then it is coherent to characterize institutions as *systems or sets of rules*. The practices we associate to an institution (e.g. exchanging goods on a market) are generated by the rule-following behaviors. Institutions and practices are the two faces of the same coin, depending on whether we focus on the rules that are followed or on the behaviors they lead to. Among the philosophers interested in social ontology, John Searle is the one whose definition of institutions is the most similar to the one defended here. However, Searle is more specific since he defines institutions as systems of *constitutive* rules [(Searle 1969, 51); (Searle 2005); (Searle 2010, 10)]. The distinction between constitutive and regulative rules is indeed central in Searle’s social ontology. Searle has grounded this distinction on two related but logically independent criteria: a metaphysical one and a syntactic one. He has introduced a third criterion only recently. The three criteria have the same implication: contrary to regulative rules, constitutive rules do not contribute to the construction of the social reality by being merely followed. While constitutive rules are constitutive of some practices (say, chess) “in the sense that playing chess is constituted *in part* by acting in accord with the rules” (Searle 1995, 28, my emphasis), they bring ‘something more’ to the social reality.

From the metaphysical point of view, regulative rules “regulate antecedently or independently existing forms of behavior” while constitutive rules “do not merely regulate, they create or define new forms of behavior” (Searle 1969, 33).<sup>3</sup> In other words, some behaviors or activities

---

<sup>2</sup> One may argue that baseball, like any other games are not really institutions but rather practices that may, but necessarily, involve institutions such as the Major League of Baseball. However, there are several reasons to reject this distinction. One of them is the fact that I can hit a homerun in a baseball game even if this game is not an official one (*i.e.* formally recognized by the MLB). Another reason is that it seems dubious to separate practices and institutions in such a way. At least in a Wittgensteinian perspective, both are indeed the two faces of the same coin.

<sup>3</sup> See also Searle (1995, 27) and Searle (2010, 9-10).

are logically impossible without the relevant constitutive rules; quite the contrary, regulative rules are logically independent of the activities they regulate. As examples of regulative rules figure the rules of etiquette or the rule stating that we should drive on the right-hand side of the road. Examples of constitutive rules from this perspective are the rules of football or chess which create the very possibility of playing those games.

The second criterion differentiates constitutive and regulative rules on the basis of their syntax. According to Searle (1969, 36), regulative rules take the form of imperatives of the kind “Do X” or “If X, then Y”. Constitutive rules are rather of the form “This X counts as Y in circumstances C”. The count-as locution is a term of specification that imposes a status on some object or entity. One of Searle’s favored examples is pieces of paper that have the status of money: “pieces of paper engraved by the Federal Reserve count as a twenty dollars bill in the United States of America”. Another example is such or such action in a chess game that counts as a bishop or a pawn move. While the syntax of regulative rules suggests that they consist in behavioral instructions that can or cannot be followed, constitutive rules take the form of statements about the status or the properties of some entities or objects. At the same time, the Y term does not only label the X term; it also embeds a set of consequences that follow from the status given to the entity or the object (Searle 1969, 36). Still, it may seem that constitutive rules cannot be followed (or *merely* followed) in the proper sense of the term. They establish a state of affairs by describing this very state of affairs.

In his later work, Searle introduces a third criterion based on the “direction of fit” of the rule that makes this last point clearer (Searle 2010, 97). Regulative rules are “standing directives” with a world-to-word direction of fit, while constitutive rules are “standing declarations” with both a world-to-word and a word-to-world directions of fit. The world-to-word direction of fit means that regulative rules have for function to affect people’s behavior in such a way that the state of affairs conform to what the rule stipulates. Constitutive rules have this direction of fit but also have the reverse one, *i.e.* they describe what the state of affairs is. Hence, constitutive rules make “something the case by representing it as being the case”. This is a sensible criterion that reinforces the point of the syntactic criterion: constitutive rules are not rules to be merely followed. The simple fact of stating the declaration and of (collectively) accepting it is sufficient to change the social reality. However, an immediate implication of these criteria and of the definition of institutions as systems of constitutive rules is that contrary to what I have assumed above, rule-following is not sufficient for the existence of institutional facts. The rest of the paper challenges this implication without completely giving up the constitutive/regulative rules distinction.

### **3. Behavioral Description and Status Description of Rules**

Searle’s distinction has been attacked numerous times, most often without an explicit recognition that there are three and not only one criterion sustaining it.<sup>4</sup> Virtually all critiques consist in pointing out either that all constitutive rules are also regulative rules or the reverse,

---

<sup>4</sup> An exception is Morin (2013).

or more radically that all rules are both constitutive and regulative. My aim in this section is to show that all critiques ultimately converge toward the same point: any rule can be described either in terms of *behavioral* properties or of *status* properties. This is sufficient to establish that contrary to what Searle's criteria may suggest, all rules are to be followed. Describing a rule in terms of behavioral properties consists in relating two or more sets of propositions about the individual's behaviors. The description in terms of status properties characterizes the rule as a relationship between two or more sets of propositions with a normative or deontic meaning (e.g. "John is a thief" and "John is punishable" or "I have the white pieces" and "I should make the first move"). Once we recognize that normative or deontic statements have behavioral implications, then it is in principle always possible to translate a status description of a rule into a behavioral description.<sup>5</sup>

Consider the metaphysical criterion first. Though it seems intuitively reasonable, it has the problem of being description-dependent. The problem lies in the fact that the decisive test for determining the nature of a rule is if the behavior in accordance with the rule could be given the same *description* whether or not the rule exists (Searle 1969, 35). Thus, the distinction is not really between two kinds of rules but rather between two kinds of description of an action or behavior. Obviously, this is not specific to action description and institutional facts. Sentences referring to brute facts like "an asteroid falls on Earth" logically imply the existence of the "constitutive objects" (the asteroid and the Earth in the present case); however, the very same facts can also be described by substituting the constitutive objects for a complete list of properties that are constitutive of these objects. Now, consider an institutional fact like "John has hit a home run in his last baseball game". The description of this fact seems to imply a reference to the rules of baseball, at least regarding what it is to hit a home run. In principle however, it is possible (though tedious) to describe the very same fact without using the terms "home run" or "baseball": we only need a statement linking a first proposition about the sufficient conditions for the sentence "John has hit a home run" to be true with a second proposition about the consequences of this state of affairs, consisting of one or several sentences that are necessarily true. The two propositions can be themselves reformulated as two sets of sentences providing a full description of all the behaviors that have taken place since the beginning of the game.<sup>6</sup>

The syntactic criterion fails on the same ground. Once again, the point is that the syntax of a rule determines only the *description* of the rule rather than its nature or properties. It only points to a linguistic distinction.<sup>7</sup> Searle (1969, 36) actually notes that all regulative rules can be reformulated as constitutive rules. However, in the case of a reformulated regulative rule "the

---

<sup>5</sup> David Lewis makes a similar point in his article "Scorekeeping in a Language Game" (Lewis 1979).

<sup>6</sup> A counterargument is mentioned by Searle (1969, p. 35-6, my emphasis): "It is possible that twenty-two men might go through the same physical movements as are gone through by two teams at a football game, but if there were no rules of football, that is no antecedently existing game of football, there is no sense in which their behavior could be *described* as playing football". In substance, this is very similar to John Rawls' (1955) "practice conception" of rules. This is a valid point, but it is perfectly compatible with the idea that all rules are to be followed. Clearly, to play football *consists in following* the rules of football. Moreover, this has nothing to do with the fact that the rules "create" or "make possible" some practices, contrary to what Searle suggests. Ultimately, this argument fails to establish Searle's distinction because it is about the proper description of some fact, not about the nature of the fact.

<sup>7</sup> This point is convincingly made by Frank Hindriks (2009). See also Guala and Hindriks (2015).

noun phrase following the ‘count as’ is used as a term of appraisal not of specification” (Searle 1969, 36). This is unpersuasive though because appraisal may require or imply some specification.<sup>8</sup> Still, even if we accept Searle’s argument, another objection is even more convincing: all constitutive rules can be reformulated into regulative rules (Hindriks 2009). This is due to the fact that any “This X counts as Y in C” rule consists in the conjunction of two separate “If A, then B” rules. Consider for instance the case of the constitutive rule for money as formulated by Searle: “Pieces of papers that have been engraved by the Federal Reserve count as money in the United States of America”. It consists into a first rule that states the properties that pieces of paper (the X term) must have to count as money in the United States of America. A second rule then specifies the properties ascribed to money (the Y term) in terms of deontic powers. The conjunction of these two rules gives:

[If pieces of paper have been engraved by the Federal Reserve, then they count as money] and  
[If something counts as money, then it has the power to buy things.]

Note that the second rule (the “status rule” as Hindriks (2009) labels it) defines money in terms of a deontic power. We can go further and obtain a simpler statement:

[If pieces of paper have been engraved by the Federal Reserve, then they have the power to buy things.]

Thus, we have *almost* reformulated a constitutive rule in terms of a “If A, then B” regulative rule. Two additional conditions are however required to complete the reformulation. First, we have to make sure that this rule is amenable to a translation into a set of propositions about behavioral facts. This is because Searle’s regulative rules take the form of (conditional) imperatives. Obviously, there is no serious difficulty to do so.<sup>9</sup> Second, as for all “if... then...” statements, the semantic content depends on the proper causal relationship between the antecedent and the consequent. Formulated as it is, the above statement looks like a conditional material. As it is well-known, conditional materials fail to capture causal relationships as they are merely equivalent to the disjunction of the conjunction of both the antecedent and the consequent and of the negation of the antecedent ( $\neg A \vee (A \wedge B)$  in standard notation). To make sure that some causal relationship holds between A and B, we must tacitly assume that the following counterfactual is true:

[If these pieces of paper had not not been engraved by the Federal Reserve, then they would not have the power to buy things.]

---

<sup>8</sup> Consider for instance the regulative rule “Drive on the right-hand side of the road”. It can be reformulated as a constitutive rule “To not drive on the right-hand side of the road counts as unauthorized conduct”. Here, it seems that the “count as” locution should be understood both in terms of appraisal and of specification: to say of something that it is “unauthorized” is to make an evaluation (an appraisal), but this evaluation implies a specification of what is for something to be unauthorized. It is simply not clear that we can separate the two because while the appraisal refers to the conduct, the specification refers to the status and both are deeply intertwined as I argue in the text.

<sup>9</sup> These behavioral facts can be described by sentences of the type “In circumstances C, if person x has given a piece of paper p to person y, then y must give back a good g to x”.

If this counterfactual is true, then we know that specific pieces of paper have the power to buy things *because* they have been engraved by the Federal Reserve. It is important to acknowledge that such counterfactual statements are actually underlying *all* regulative rules that have the “if... then...” form. For instance, the regulative rule “if the light is red, then stop” is indeed a rule if and only if the fact that the light is red matters in the fact that you should stop. Hence, the counterfactual “if the light were green, then you would not have to stop” must also be true.

The point is thus that the syntactic criterion grounds the distinction between constitutive rules and regulative rules on a mere linguistic convenience: if we want to refer to a rule on the basis of a status description, the “X counts as Y in C” form is more straightforward and less ambiguous; if we want to refer to the same rule on the basis of a behavior description, the “if A, then B” form is more natural. Ultimately however, all rules can be formulated in this latter way. This is important because it indicates that we can speak of rule-following behavior for all kinds of rules.

#### 4. The “Direction-of-Fit” Criterion for Constitutive Rules

As I have noted in the preceding section, Searle (2010, 97) has introduced a new criterion to sustain his distinction. Let me call it the “direction-of-fit criterion”. In this perspective, constitutive rules are “standing declarations” which have both the world-to-word and the word-to-world directions of fit, while regulative rules are “standing directives” which only have the former one. As far as I can tell, this criterion evades the objections made above because it does not depend on the way we describe rules (in terms of status or in terms of behavior).<sup>10</sup> However, on a rule-following account, it seems that all rules have the double direction of fit. This is due to the fact that intentional rule-following implies that the rule is self-fulfilling and self-reflexive.<sup>11</sup> Intentional rule-following means that a rule is followed on the basis of an intentional attitude (or a set of intentional attitudes) related to the subjectivity of the intentional being (Pettit 1990). This implies that the rule (whatever its description) itself must be part of the propositional content of the relevant attitude.

A minimally rational intentional being must form attitudes that verify conditions of satisfaction defined by their direction of fit.<sup>12</sup> In particular, an intention to follow a rule is necessarily causally self-referential: the satisfaction condition of the intention corresponds to its propositional content. In other words, the intention to follow the rule succeeds if and only if the intention causes actual rule-following behavior. Generally, this intention will have a cognitive counterpart corresponding to one or several beliefs. Ignoring the controversial case of private rules, rule-following necessitates for each person to form a belief about whether or not others

---

<sup>10</sup> Note however that while the term “directive” more naturally applies to the behavioral description, the term “declaration” is easier to understand in terms of status description.

<sup>11</sup> See Bloor (1997, 27) who develops a “self-reflexive model” of institutions on the basis of Ludwig Wittgenstein’s account of rule-following [(Wittgenstein 1965); (Wittgenstein 1981); (Wittgenstein 2010)]. Bloor’s definition of self-reflexivity is stronger than mine however since he requires that an institution or a practice cannot be defined without referring to the practice (its rules) itself. I disagree with this but the point is that Bloor seems to hold that *all* rules are self-reflexive in this way.

<sup>12</sup> Searle (2010, 38-41) provides a useful summary of the different intentional attitudes and their direction(s) of fit.

are also following the rule. The propositional content of a belief is satisfied if it conforms to how the world actually is (*i.e.* beliefs have a word-to-world direction of fit), which in this case means that others are actually following the rule. If everyone's belief about the fact that others are actually following the rule is true and if everyone succeeds in forming the intention to following the rule, then it is a tautology that everyone actually follows the rule. Rule-following thus entails at least a belief and an intention. Since both attitudes have opposite directions of fit, *all* rules function on the basis of the double direction of fit.

## 5. Alternative Accounts of Constitutive Rules

Should we completely abandon the idea that some rules are constitutive? Is the notion of constitutive rules completely incompatible with a rule-following account of institutions and social ontology? I shall argue for a negative answer to both questions in this section and the next one. The notion of constitutive rules relies on an intuition that some rules are more fundamental, “thicker” than others. However, this intuition does not determine the criterion of “thickness” of rules and Searle's three criteria discussed above do not fully exhaust all the possibilities. Other distinctions between kinds of rules have been proposed and are compatible with the notion of rule-following in spite of the fact that they are often discussed (and sometimes assimilated to) along with Searle's constitutive rules. Ludwig Wittgenstein's later writings on language games and rule-following distinguish between several kinds of rules on the basis of their status relatively to a given practice. This is particularly the case in *The Blue and Brown Books* (Wittgenstein 1965) where he suggests that a rule can be either “an instrument used in what we should call *the practice* of the game” or “an instrument for the *training*” in playing the game (Wittgenstein 1965, 96, emphasis in original). Call this the *Practice Case* and the *Learning Case* of rules respectively. Wittgenstein (1965, 98-9) also notes that some rules correspond to neither case and rather correspond to *descriptions* of the behavior of some group of persons. We can call this the *Summary Case* of rules. It should be clear that these three cases are easily understandable in terms of rule-following: in the Practice Case and the Learning Case, rule-following consist in conforming to the instructions given by the rule; in the Summary Case, the rule and the behaviors are conflated.<sup>13</sup>

A similar account is given by Rawls (1955). Rawls argues that we should not conflate the fact of justifying a practice with the fact of justifying a particular activity falling under a practice. The former is amenable to utilitarian considerations: for instance, the practice (institution) of promise keeping is clearly useful to help people to achieve a better satisfaction of their wants and desires because it fosters cooperation in a population.<sup>14</sup> This is not the case for the latter because the activity is bounded and even determined by the practice. In the case of promise keeping, the fact of keeping his promise in a particular instance is not determined by the utilitarian principle (*i.e.* it is in the interest of the promisor to keep his promise) but rather

---

<sup>13</sup> Of course, I ignore here all the difficulties related to Wittgenstein's account of rule-following and whether or not he considered that there is a fact for following a rule. A well-known skeptical interpretation is offered by Kripke (1982). Bloor (1997) and Pettit (1990) develop on the contrary positive answers.

<sup>14</sup> This can be accounted for on the basis of an evolutionary game-theoretic approach. See Hédoin (2015) and Hindriks and Guala (2015) for further discussions on the different accounts of rules in a game-theoretic framework.

because *it is part of the practice of promise-keeping*. This directly leads to two corresponding conceptions of rules. On the *summary view*, “rules are pictured as summaries of past decisions arrived at by the *direct* application of the utilitarian principle to particular cases” (Rawls 1955, 19, emphasis in original). Rules are thus descriptions of salient behavioral regularities that have emerged on the basis of the individuals’ practical and epistemic rationality. Quite differently, according to the *practice conception* “rules are pictured as defining a practice... rules define a practice and are themselves the subject of the utilitarian principle” (Rawls 1955, 24). On the practice conception, it is insufficient that individuals act *as if* they were obeying the rules; rules must be actually followed as part of a practice. This implies that rules must be “publicly known and understood as definitive” (Rawls 1955, 24).

It is worth making a comparison between Wittgenstein’s and Rawls’ respective accounts and to put them in perspectives with Searle’s metaphysical criterion. First, there seems to be a deep affinity between Wittgenstein’s Summary Case and Practice Case on the one hand, and Rawls’ summary view and practice conception, on the other hand. Moreover, in the case of the practice conception, Rawls emphasizes the importance of the learning and teaching of rules: “It is the mark of a practice that being taught how to engage in it involves being instructed in the rules which define it, and that appeal is made to those rules to correct the behavior of those engaged in it” (Rawls 1955, 24). This makes it closely related to Wittgenstein’s Learning Case. There are also similarities between Rawls’ practice conception of rules and Searle’s metaphysical criterion for constitutive rules. On the practice conception, the practice is logically prior to particular cases in the sense that “given any rule which specifies a form of action (a move), a particular action which would be taken as falling under this rule given that there is the practice would not be *described as* that sort of action unless there was the practice... No matter what a person did, what he did would not be described as playing baseball, and for him to be doing this presupposes the rule-like practice which constitutes the game” (Rawls 1955, 25, my emphasis). Searle makes a very similar statement:

“It is possible that twenty-two men might go through the same physical movements as are gone through by two teams at a football game, but if there were no rules of football, that is no antecedently existing game of football, there is no sense in which their behavior could be *described as* playing football.” (Searle 1969, 35-6, my emphasis).

As I note above (see fn. 6), this characterization of constitutive rules (or rules of practices) is not in contradiction with the notion of rule-following. However, it fails to establish that there are two *kinds* of rules because whether or not a rule is “logically” a particular case depends on the appropriate *description* of the practice. Note however that this does not make Wittgenstein’s and Rawls’ accounts irrelevant because the summary view of rules does not really deal with rules but rather with behavioral patterns. Therefore it is still possible (and correct) to distinguish between behavioral patterns that can be *described as if* they were the result of rule-following and rules that are *instruments* for a practice and lead to *intentional* rule-following behaviors. Searle’s regulative and constitutive rules are both of this latter kind.

## 6. Constitutive Rules Again: Practical Rationality and Validity of Intentional Attitudes

Wittgenstein's and Rawls' practice conception of rules include both regulative and constitutive rules in Searle's sense. Under some description of a given practice, any rule can be conceived as an instrument used in that practice. I believe that this tends to indicate that to seek the distinction between constitutive and regulative rules in the *ontology* of rules is hopeless. However, I want to argue that an alternative way to ground this distinction is available while fully in accordance with a rule-following account and the practice conception.

As indicated in the preceding section, the intuitive idea underlying this distinction constitutive and regulative rules is that some rules are more essential than others for a given practice. Of course, if a practice is co-extensionally defined by the corresponding set of rules then all rules are essential. However, this is once again a matter of definition or description and this is not – I contend – what the intuition is about. My suggestion however is that whether or not a rule is essential (and so, “constitutive” of the practice in this particular sense) is not a matter of ontologically objective facts but rather of ontologically subjective but *epistemically* objective ones. This is suggested in an allusive way by Wittgenstein (2010, §564) who noted that that the distinction between essential and inessential rules depends on the *point* of the game. To the remark that a game is defined by its rules and that all rules are essential in this definitional sense, he suggested the objection that “one does not see the point of [the rule's] prescription” (Wittgenstein 2010, §567). In other words, the constitutive nature of a rule depends not on the fact that it is part of the definition or the description of the practice, but on the fact that it has a peculiar relevance given the whole point of the practice.

What determines the point of a practice and, on this basis, the relevance of a rule? There are several possibilities here. According to a first approach, the point of the practice is basically determined by the formal consistency of the set of rules  $S$  and the essentiality of each rule is evaluated against this formal consistency. By formal consistency, I mean here the fact that in general an institutional practice should not depend on two incompatible rules, i.e. rules leading to conflicting behavioral prescriptions.<sup>15</sup> Assume that the set  $S$  can be divided into two subsets: a set  $S'$  of rules which cannot be changed/suppressed without affecting the formal consistency of  $S$  and a set  $S''$  of rules which do not have this property. Then, it might be argued that only the rules in  $S'$  are constitutive of the corresponding practice because they cannot be changed without making the practice wholly inconsistent. There are several problems with this approach however. A first one is that whether a rule belongs to  $S'$  or  $S''$  may depend on which new rule will take its place.<sup>16</sup> A second problem is that even though a given rule may belong to  $S''$ , it

---

<sup>15</sup> The simplest case is the one where two rules  $R$  and  $R'$  have the same antecedents but different consequents. But in sufficiently complex practices, there may be pairs of inconsistent rules even with different antecedents. This will be the case for instance if the change in the consequent of a rule  $R'$  makes the conjunction of the antecedent and the consequent of a rule  $R$  impossible but the antecedents of both rules are correlated in the practice.

<sup>16</sup> The rule may also be simply suppressed rather than changed.

may indirectly have an impact on the partition of  $S$ .<sup>17</sup> As a consequence, the essentiality of a rule is always relative to some given set of rules  $S$ .

Though these problems do not necessarily disqualify this first approach, they make it less appealing. A more attractive alternative is to point out that the essentiality of a rule may not depend on its *formal* consistency with the whole practice but on what I will call its *intentional consistency*. This point can be made by using the remarks made in section 4 about the role of intentional attitudes in rule-following. If we ignore the case of private rules, rule-following implies at least two kinds of intentional attitudes: a belief that others are following the rule and an intention to follow the rule. The former has a word-to-world direction of fit (*i.e.* the propositional content of the belief must match with the actual states of affairs) and the latter a world-to-word direction of fit (*i.e.* the intention is satisfied if it brings an action that leads to a state of affairs that conforms to the propositional content of the intention). Successful rule-following behavior thus supposes that the individuals form the appropriate intentional attitudes. Most of the time however, a rule will not be followed in isolation but as part of a whole institutional practice against which the *point* (to use Wittgenstein's word) of the rule will be evaluated. This evaluation is possible because all institutional practices have a *value*; more exactly, when we engage in some institutional practice we do so in part because we ascribe to this practice a value.<sup>18</sup> This is pretty much evident for games like chess or baseball where the value is derived from the fact of defining some goals and voluntarily establishing and submitting to obstacles that have to be overcome to reach these goals (Suits 2005). This is also the case for any other practice like getting married, making monetary exchanges, voting to an election, and so on. There are reasons why we engage in these practices and follow the corresponding rules and these reasons find at least partially their origins in the value of the practice.<sup>19</sup>

It seems reasonable to assume that the value of a practice counterfactually depends on the rules that define it. Call it the "Value-Rule Counterfactual Dependence Claim":

*Value-Rule Counterfactual Dependence Claim* – For any institutional practice  $P$  defined by a set of rules  $S$ , the value  $V$  of  $P$  is counterfactually dependent on  $S$ , *i.e.* some changes in  $S$  would have implied a change in  $V$ .

Another way to put this claim is that "engaging in the practice constituted by  $S$  is valuable (at least for those who engage in it) in ways in which it *could not have been valuable without the existence of  $S$* " (Marmor 2009, 39, emphasis in original). More exactly, the practice  $P$  would have had a different value  $V$  (and maybe no value at all) if  $S$  had been different. The *Value-Rule Counterfactual Dependence Claim* does not imply that value reduces to the rules defining a practice because of multiple realizability: different sets  $S$  can realize the same value  $V$ . This is

---

<sup>17</sup> Denote  $S''^*$  the set  $S''$  of inessential rules after a change in one of its components. It is perfectly possible that the partition of  $S$  is affected by this change, *i.e.* that rules that belonged to  $S''$  before the change now belongs to  $S''^*$ . The converse is also true, obviously.

<sup>18</sup> See also Marmor (2009, 36-44).

<sup>19</sup> There may be other reasons akin to Rawls's utilitarian principle, *i.e.* personal interests. That means that rule-following is clearly partially instrumental but it is hard to argue that we can reduce any practice as a whole to such kinds of interests.

indeed a key point here: some rules may change or be suppressed without changing the value of a given institutional practice. These rules may be said to be “inessential” to the value of the practice. Other rules however will be regarded as essential because they cannot be changed or suppressed without affecting the value of the practice. Consider the example of chess: rules regarding the time allowed to each player to play their moves may be regarded as inessential to the practice; to the contrary, rules related to the movement of pieces should probably be regarded as fundamental for the value of the game. There are also borderline cases such as for the rules regarding the initial disposition of the pieces at the beginning of a match.<sup>20</sup>

The example of chess highlights an important point regarding the nature of the separating line between essential and inessential rules. From a strict ontological point of view, this line cannot be determined objectively. As I have emphasized in the preceding sections, all the attempts to distinguish constitutive from regulative rules in this way fall short. The very fact of grounding the distinction on the notion of *value* and on the *Value-Rule Counterfactual Dependence Claim* indicates that it must be ontologically subjective: value does not exist independently of the persons who ascribe it to something. However, this does not mean that “anything goes”: the essentiality of some rules is reflected in the person’s attitudes toward the practices and their rules. Value is not purely subjective but rather inter-subjective: since we are ignoring the particular case of private rules, institutional practices are by definition collective practices common to the members of some community. In some way, the very fact of being the member of a community *is to* participate in some practices along shared public rules. The value of the practice thus does not depend on one individual in particular but on the whole community. This implies that the status of a given rule (“constitutive”, “essential”) is a function of the intentional attitudes of all the members of the community.<sup>21</sup>

How this translates at the level of each individual’s intentional attitudes? I have already explained why intentional rule-following implies for each individual to form a belief regarding the fact that others are also following the rule, and an intention to follow the rule. However, rule-following is derived from the fact of participating to a practice, *i.e.* we never follow a rule in isolation but in connection with other rules. The significance of a rule thus depends on the other rules that surround it as parts of the practice. Thus, I submit that intentional rule-following properly speaking implies that the propositional content of the intentional states (beliefs and intentions) includes the whole practice and (at least implicitly) its value. More formally, we can derive the two following formulae:

*Belief*: [The members of the collective *C* are following the rule *R* as part of the practice *P* with value *V*]; *c.s.*: the members of *C* actually follow *R* as part of *P* with value *V*.

---

<sup>20</sup> There is a variant of chess called “random chess” where the initial disposition of the pieces is determined randomly before each match. This variant has been played a few times in official competition by professional chess players.

<sup>21</sup> As I will make it clear below, I am not assuming that this necessarily depends on the existence of *collective* intentional states. Whether or not there are intentional states that do not reduce to individual intentionality is an issue that is not essential to my point.

*Intention*: [I follow  $R$  as part of  $P$  with value  $V$  by performing the required action]; *c.s.*: this intention cause: I follow  $R$  as part of  $P$  with value  $V$  by performing the required action.

The first term in italics in each formula refers to the kind of intentional attitude. The sentence into brackets refers to the propositional content. The second sentence specifies the conditions of satisfaction *c.s.* for the intentional attitude to be satisfied. The self-referential nature of intentions is reflected by the fact that the propositional content is causally referred to in the conditions of satisfaction, which is not the case for beliefs. Once again, it should be emphasized that how the rule  $R$  and the practice  $P$  are described is irrelevant.

On this basis, I claim that the constitutive or essential nature of a rule is derived by imposing on each intentional attitude a *condition of validity*. This condition of validity should indicate if the propositional content of the formula is meaningful. More exactly, it has to state if the “rule  $R$  as part of the practice  $P$  with value  $V$ ” part of the propositional content as any sense at all. How this is figured out depends on the value  $V$  of the practice  $P$  and on the resulting significance of the rule  $R$ . As I said above, for public rules and collective practices the value of a practice and the significance of a rule is neither objectively given nor purely subjective. It is ontologically subjective but epistemically objective because it is imposed by the whole community. Consider again the example of chess. When expecting to play a non-official match against someone, there are some rules upon which you can argue, *e.g.* the rule regarding the time given to the players to play their moves. For other rules, you cannot even imagine submit a change to the other players because you expect a response of the kind “well, actually you do not want to play chess” or “this is not how we play chess”. Actually, you would probably not be able to play any match because nobody would want to take part into the new practice you are creating. This is simply due to the fact that the rule changes you are submitting enter into conflict with the value of the practice as it is conceived by the community of players. As members of the community of chess players, everybody expects everyone else to recognize that some rules are indeed essential to the practice. This is part of the practical rationality of the chess player.<sup>22</sup>

This idea can be stated more generally and more formally. The practical rationality of the members of a given collective or community  $C$  taking part in a practice  $P$  sets the conditions of validity  $CV$  upon the meaningful intentional states. We can denote  $CV(Beliefs)$  and  $CV(Intentions)$  respectively the sets of all the valid beliefs and intentions that the members of  $C$  can have when participating in  $P$ . Not all the rules that are part of the propositional content of these valid intentional attitudes are “constitutive” or “essential” rules: some of them can be substituted or eliminated validly, *i.e.* the new rule figures in the propositional content of an attitude belonging to  $CV(Beliefs)$  or  $CV(Intentions)$ .<sup>23</sup> It is possible that there are rules for which

---

<sup>22</sup> There are some affinities with Wittgenstein’s well-known concept of *lebensform* (“forms of life”), which is basically a tacit agreement over what it implies to be part of some practice. It is not clear this is totally equivalent though, since Wittgenstein used this concept as a way to put an end to the infinite regress of interpreting a rule with another rule.

<sup>23</sup> In the case where the rule is simply eliminated, the criteria would be whether or not the intentional attitudes referring to the remaining rules corresponding to the practice remain valid.

there are no alternative rules with valid intentional attitudes.<sup>24</sup> Such rules would be “constitutive” of or “essential” to the practice as conceived by the members of the community since they could not be given up for another rule to be intentionally followed as part of the practice. The content of the sets  $CV(\text{Beliefs})$  and  $CV(\text{Intentions})$  for a given practice depends on the individuals’ practical rationality. But the practical rationality of someone participating in a practice is partially determined by the very fact of being a member of the community of the people participating in the practice. As a consequence, the constitutive nature of a rule is not something to be determined purely subjectively by each person. It is a feature that is derived from the fact of being the members of some community.

## 7. Tentative Answers to Possible Objections and Conclusion

I have developed in this paper an account of “constitutive” or “essential” rules that is not only compatible with the notion of intentional rule-following but is also immune to the problems of the standard accounts of constitutive rules, in particular Searle’s one. The main line of the argument is that Searle’s various criteria to distinguish constitutive rules fail to establish the distinction with regulative rules because they depend on how rules and practices are described. This is not surprising because ultimately, all rules have the same function and work the same way. Other accounts of rules developed for instance by Rawls and Wittgenstein develop a distinction between “as-if” rule-following behavior and rules used as instruments in a practice. Though their “practice conception” of rule does not distinguish between essential and non-essential rules, it points to an important relationship between rules and practices.

My own account of constitutive or essential rules depends on several claims: 1) that rule-following requires one to have the relevant intentional attitudes (beliefs and intentions), 2) that each practice has a value which is ascribed by the members of the community who participate in it, 3) that value counterfactually depends on rules (the *Value-Rule Counterfactual Dependence Claim*). Combined, these three claims entail that the essentiality of a rule is ontologically subjective but epistemically objective: essential or constitutive rules are rules that cannot be substituted by other rules followed with a valid intentional attitude. This is due to the fact that some rules are simply incompatible with a practice given the value given to it by the participants.

There are several more or less obvious objections that can be made to this new account of constitutive rules. I discuss and answer the most significant ones in this concluding section. I do not intend to provide definitive answers, but this should be sufficient to show that my account is comparatively less rather than more problematic than the Searlian constitutive/regulative rules distinction.

---

<sup>24</sup> Conditions of validity can be formulated in terms of a possible worlds semantics with an accessibility relation. A rule is said to be constitutive or essential in actual world  $w$  if there is no  $w$ -accessible worlds with an alternative rule in the propositional content of the individuals’ intentional attitudes. The accessibility relation should be defined in this case such that all  $w$ -accessible worlds instantiate the same set of values as world  $w$ . Note that because of multiple realizability, this does not imply that these worlds all have the same set of rules.

*Objection 1: The value of a practice may be unrelated to its rules.*

This is a direct objection to my *Value-Rule Counterfactual Dependence Claim*. It can be substantiated in the following way: suppose that initially I started to play chess because I found this game (and thus its rules) enjoyable and intellectually challenging. As time goes by, this has been less and less the case and today I play chess essentially because it is my principle source of monetary income (e.g. I am a professional chess player). Moreover, my valuation of the practice is independent of the changes in its rules that may have occurred during the relevant period. This is quite general: we may participate in a practice and thus value it for many reasons; the value of the practice comes from the reason for following its rules rather than its rules *per se*. Call it the “*Reason-Based Value Claim*”.

Answer: The *Reason-Based Value Claim* is highly plausible but is not incompatible with the *Value-Rule Counterfactual Dependence Claim*. In the example above, I continue to value playing chess. While my reasons for valuing to play chess have changed, what matters is that I still accord sufficient value to this practice to continue participating in it. Now, suppose that the rules of chess are (drastically) changed in such a way that I no longer have the competences to be a good player and to earn a sufficient income. The change of rules has affected the value of the practice. Of course, a decline in my intellectual capacities may directly affect the value of the practice without changing the rules of chess. But this is not in contradiction with the *Value-Rule Counterfactual Dependence Claim*. All in all, it seems hard to find cases where the value of a practice is *completely* unrelated to its rules.

*Objection 2: There is no rule that cannot change without affecting the value of a practice.*

Here, the claim is that the dependence between any rule and value is never systematic: there is no example of rules that are constitutive or essential in such the way that a change in them would *necessarily* affect the value of the corresponding practice. Therefore, the whole approach proposed above to distinguish constitutive from regulative rules is misguided.

Answer: I think this objection can be answered through two related arguments. First, rules scarcely if ever work in isolation from other rules. This is precisely the point of thinking in terms of *counterfactual* dependence: counterfactual reasoning should be generally done in such a way that a change in the relevant variable needs to be completed by a change in the other causally related variables. That means that while we may temporarily accept that the change of one rule can never systematically affect the value of a practice, we nevertheless should account for the fact that such a change will generally imply other rule changes. Second, as I have emphasized above, the dependence of value on rules is ontologically subjective. The truth value of the objection thus depends on the participants’ subjectivity and I would conjecture that some rules are taken as essential precisely because a change in them is subjectively expected to foster additional rule changes.

*Objection 3: There is no reason to think that everyone will agree on the value of a practice.*

This objection is somehow related to the first one above. As pointed out then, we may participate in a practice and value it for many different reasons. The identification of constitutive rules may be impossible (i.e. there may be no constitutive rules) in cases where people participate in a practice for completely different and conflicting/incompatible reasons.

Answer: We cannot exclude in principle the theoretical case where the participants in a practice have such heterogenous and conflicting reasons. But it is hard to think of concrete examples of such case. This is not really surprising: it is hard to see how this kind of practice could be sustained on a regular and durable basis. In a game-theoretic perspective, rules should be in equilibrium (Guala and Hindriks 2015), which means that there must be a minimal agreement over the value of the practice. This is reflected in the belief component of rule-following. For example, opponents in a sports competition or in game, though they have conflicting interests, must somehow agree on following rules. In practice, we may even expect people to group in communities satisfying the required minimal homogeneity (e.g. communities of professional and amateur players).

*Objection 4: This account is unable to reflect on (pervasive) diachronic disagreements over the value of a practice.*

This is the diachronic version of the preceding objection. As a matter of fact, it is pretty clear that the value given to a practice may change as time unfolds without any change in its rules. For instance, we may value marriage for very different reasons than two centuries ago, while the rules of marriage may not have (significantly) changed. Moreover, some rules may be regarded as constitutive (resp. regulative) at some time, but regulative (resp. constitutive) at others.

Answer: This is the most challenging objection. A tentative way to answer it involves most of the arguments used to respond to the preceding objections: 1) the *Value-Rule Counterfactual Dependence Claim* does not imply that a change in the value of a practice is impossible without a change in its rules, 2) the status of a rule (constitutive or regulative) is ontologically subjective and may thus change through time, 3) some *dynamic* agreement over the value of a practice is required to account for its continuity as time unfolds. This last point suggests that what matters for the identification of an institution through time is not (or not only) its core set of constitutive rules but rather its *phylogeny*.

*Objection 5: This account does not escape some form of essentialism, especially regarding the status of communities.*

This account may be labeled “essentialist”. This may be perceived as a weakness because it seems to indicate that it relies on too strong assumptions regarding our intuitions about what is essential to an institution. The essentialist criterion is extremely demanding because “[f]irst, it

requires us to make predictions about future instances of the institution, and about some present and past instances that we don't know about, but might discover some day" and "[s]econd, many social practices lack a stable set of core social features" (Morin 2013, 9). In particular, the condition of universal agreement inside a given community is required to identify the constitutive rules, which implies to essentialize communities.

Answer: These are fair points but my version of the essentialist criterion is immune to them, at least as far as practices are concerned. The reason is that it makes absolutely no claim regarding the nature of institutional practices and their core features. It merely points out the fact that it is generally agreed in a given community what are the key rules that have to be followed because of the value of the practice. This translates at the behavioral level where there may be strong resistance to the change in the rules. Thus, a given practice will retain core rules as time unfolds, without anyone presupposing that this will be eternally the case. However, my account *does* indeed rely on some essentialist view of communities. I do not see any obvious way to remedy to this problem and it probably comes with the Wittgensteinian understanding of rule-following. If this is considered as a weakness, then this should be put in perspective with the problems of Searle's account of constitutive rules discussed in the first half of the paper. The reader is left as the judge to determine which approach is nevertheless the most satisfactory.

## References

- Bloor, David. 1997. *Wittgenstein, Rules and Institutions*. Routledge.
- Epstein, Brian. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford University Press, Incorporated.
- Guala, Francesco, and Frank Hindriks. 2015. "A Unified Social Ontology." *The Philosophical Quarterly* 65 (259): 177–201.
- 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 (2): 253–75.
- Hindriks, Frank, and Francesco Guala. 2015. "Institutions, Rules, and Equilibria: A Unified Theory." *Journal of Institutional Economics* 11 (03): 459–80.
- Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Harvard University Press.
- Lewis, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8 (1): 339–59.
- Marmor, Andrei. 2009. *Social Conventions: From Language to Law*. Princeton University Press.
- Morin, Olivier. 2013. "Three ways of misunderstanding the power of rules." In M. Schmitz, B. Kobow and H-B. Schmidt (eds.), *The Background of Social Reality*, 185-201.
- Pettit, Philip. 1990. "The Reality of Rule-Following." *Mind*, New Series, 99 (393): 1–21.
- Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64 (1): 3.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

- . 1995. *The Construction of Social Reality*. Simon and Schuster.
- . 2005. “What Is an Institution?” *Journal of Institutional Economics* 1 (01): 1–22.
- . 2010. *Making the Social World: The Structure of Human Civilization*. Oxford University Press.
- Suits, Bernard. 2005. *The Grasshopper: Games, Life and Utopia*. Broadview Press.
- Wittgenstein, Ludwig. 1965. *The Blue and Brown Books*. HarperCollins.
- . 1981. *Remarks on the Foundations of Mathematics*. Blackwell.
- . 2010. *Philosophical Investigations*. John Wiley & Sons.