The Institutionality Of Legal Validity

KENNETH M. EHRENBERG

University of Surrey

The most influential theory of law in current analytic legal philosophy is legal positivism, which generally understands law to be a kind of institution. The most influential theory of institutions in current analytic social philosophy is that of John Searle. One would hope that the two theories are compatible, and in many ways they certainly are. But one incompatibility that still needs ironing out involves the relation of the social rule that undergirds the validity of any legal system (H.L.A. Hart’s rule of recognition) to Searle’s notion of codification: the idea that institutions need official declarations of their constitutive rules in order to enjoy the full benefits of institutions. The incompatibility arises from the fact that, in order to do its institutional work, the basic validity rule must be codified in Searle’s sense—yet, given the particular role it has in legal positivism, it may be impossible to codify in the Searlean sense. In this paper I develop the incompatibility in detail, consider and reject consigning the basic validity rule to Searle’s “Background” capacities that support institutional facts, and conclude that the best route to eliminating it while doing a minimum of damage to the two theories is to make a slight emendation to Searle’s theory of institutions.

The most influential theory of law in current analytic legal philosophy is that of H.L.A. Hart (1994). Hart and his followers understand law to be a kind of institution. The most influential theory of institutions in current analytic social philosophy is that of John Searle (1995, 2010). Hence, one would hope that the two theories are compatible in...
what they tell us. In many ways they certainly are. But some incompatibilities still need ironing out.

One key feature of institutions is their ability to impose norms independently of the beliefs and intentions of those subject to the norms (that is the very point of institutions) (Searle 2010, 23, 85-86; Shapiro 2011, 210-11). Institutions are capable of creating new reasons for action, perhaps even reasons independent of actors’ interests. By manipulating incentives through sanctions and the possibility of public disapprobation, the law exemplifies this by acting upon our subjective or motivating reasons.

But institutions also go further in purporting to impose new normative or objective reasons on institutional members. The law is again a prime example. Legal institutions like corporations seem to give rise to new reasons for action, perhaps even obligations to act against our interests, even if we are otherwise reluctant to accede to any general obligation to obey the law. A complete account of law, or of institutions more generally, then owes us an explanation of how facts about certain special kinds of social interactions give rise to these new norms.

One way of understanding Searle’s theory of institutional facts is as an attempt to give such an explanation of how facts about institutions might create new norms. If successful, this would be particularly valuable to legal positivists who follow Hart, since by seeing law as a species of social fact, they are in need of an explanation of how to get from fact to norm. In order for Searle’s theory to be useful in that way for legal positivists, however, we would have to make sure what Searle tells us about institutions is consistent with what legal positivists tell us about the law. This paper is an initial step in that project by serving as an attempt to iron out one point of tension.

A feature of legal systems, often noted by philosophers of law, is that they involve norms that pick out which particular laws are legally valid under that legal system. A law of one jurisdiction, is not necessarily legally valid in another, where the two jurisdictions have different legal systems with different rules to tell us how to identify which laws are valid. Similarly, if the legislature of one of these jurisdictions tried to pass a law that was not in conformity with its own rules of validity, the putative law would not (generally) be recognized as legally valid within that jurisdiction. Even if a law is identically worded to one in another legal system (as frequently happens with uniform codes adopted across multiple states in the U.S.), it needs to be adopted according to the procedural rules of each jurisdiction to be valid in that particular jurisdiction. Of the legal theories that treat issues of legal validity, legal positivism goes furthest in its emphasis on the socially constructed—institutional nature of law (and

---

4 An extended analysis of their points of compatibility can be found in Ehrenberg (2016).
5 The possibility of any normative demands being based on mere (non-normative) facts is, of course, a huge matter of contention in the philosophy of value, with Hume claiming this is impossible. It is a question on which Searle has been attempting to shed light for quite some time. See Searle (1964). I make no comment on the success of those attempts. But even if normative demands cannot be based merely on descriptive facts, facts are certainly relevant to directing them.
6 Neil MacCormick perhaps focused the most on the relation between theories of institutional facts like Searle’s and of legal positivism such as Hart’s. Yet it is not clear that he appreciated the particular difficulty raised in this paper as he did not consider the law itself to be an institution, considering instead legal entities such as contracts and wills to be institutions of law (MacCormick, 1974, 105).
7 See Schauer (2005, 496).
most aspires to an explanatory jurisprudence), again owing us an account of the gene-
sis of legal normativity.  

Contemporary legal positivists have argued that legal systems rely on some basic rule, 
which settles the most basic questions of a jurisdiction’s validity conditions, a rule that 
contains the most basic criteria of legality for that jurisdiction. For Hart and others who 
follow his form of legal positivism, this is the “rule of recognition.” This basic rule is 
what legal officials follow when they uphold their jurisdiction’s most fundamental laws 
and procedures in the application of existing laws and the generation of new ones. 
According to Hart and those who follow his general lead, collective acceptance and 
application of this basic validity rule by key officials determines what counts as legally 
binding within a given jurisdiction. Hence this rule helps to determine the content and 
application of the jurisdiction’s legal norms, which are key institutional facts about the 
legal system. The question, which Searle’s theory holds out the promise of answering, is 
how to get from the fact of collective acceptance of the jurisdiction’s basic validity rule 
on the part of key officials, to the normativity of law—its apparently binding character 
on the rest of us. This paper is aimed at aiding in that project by examining a specific

---

8 David Enoch has argued that there is no problem in accounting for the ability of law to create normative reasons as it merely triggers pre-existing conditional reasons in basically the same way that a grocer raising the price of milk gives you a reason to buy less—by manipulating non-normative facts (2011, 4-5, 26-28, his example). One issue I take with his analysis is that it depends upon seeing contextually-bound normative reasons as “imaginary” if they do not pass a moral test, as only morality gives “real” normative reasons. Ibid., claiming “machismo-honor practice” does not give one real reasons. It seems to me that when one is engaged in a practice, one might get contextual reasons relating to that practice (reasons from the point of view of the practice, an institution would be a prime example), even if the immorality of that practice ultimately means those reasons are easily outweighed or vitiated. I wish to deny that these contextual reasons are “imaginary,” although I don’t think that doing so commits me to a specific meta-physical position about them or what is meant by calling them “real.” Enoch’s view would not allow a distinction between a mistake made about the reasons the practice is imposing (e.g., that the machismo-honor practice requires severe tickling of those who dishonor you) and a mistake made in what morality ultimately requires. It also has the implication (unpalatable to me) that outweighed reasons cease to be “real” reasons rather than reasons that no longer apply. Veronica Rodriguez-Blanco also points out (perhaps even more germane to our discussion here) that Enoch’s description of the way that non-normative facts trigger pre-existing conditional reasons still leaves mysterious the move from fact to triggered norm (Rodriguez-Blanco, 2013, 22). Finally, even if all legal reason-giving is a triggering of pre-existing conditional reasons, it would be nice to have an account of what might make plausible law’s apparent claim to give reasons even to those justifiably not engaged in being law-abiding (assuming that those already engaged in the project of being law-abiding get their reasons triggered as Enoch describes, and those not justifiably disengaged from that project already have a “real” moral reason to engage with it).

9 Hart (1994, 94-110). For Hans Kelsen, this is the Grundnorm. (1967, 8). There are some differences between Hart and Kelsen that are not immediately relevant here (for a quick analysis see Waldron, 2009, 332). In my exposition below I tend to follow Hart over Kelsen (although I will generally use the more general and descriptive “basic validity rule” rather than the theory-specific and more technical “rule of recognition”). Some of the reasoning for preferring Hart’s analysis to Kelsen’s is given by Marmor (2006a, 348-53). Joseph Raz has argued that there can be multiple rules of recognition within a given legal system (1975, 146-48). Scott Shapiro has argued that “the rule of recognition should be identified with all of the norm-creating and norm-applying parts of [a] shared plan” of constitutional order (2009, 250, emphasis in original). One result of this view is that only a portion of the rule of recognition is considered ultimate. Ibid. The problems raised below apply just as easily to such possibilities, so I will write as though there is a unitary basic validity rule for simplicity.

10 For another attempt to answer this question that does not focus as much on law’s institutional nature, see Himma (2013).
difficulty in harmonizing Searle’s theory with legal positivism and suggesting a possible solution.

Searle relies upon the notion of “codification” to explain how the validity of institutional norms gets specified. That is, he uses it to explain what counts as a robust instance of an institutional norm in formal institutions. The problem is that the basic validity rule cannot be understood as a codified rule in Searle’s sense, yet it seems that it must be understood as codified in order for it to do the work that legal positivists have given it. This paper will develop the problem in some detail and conclude by suggesting a modification to Searle’s theory in order to harmonize the theories.

I begin with a sketch of Searle’s theory of institutional facts, just to situate his notion of codification within it. Then I move on to show the relationship between a legal system’s basic validity rule and collective intentionality. Then I show how that interplay creates problems for the work that codification is supposed to do in the creation of institutional facts, especially in relation to deontic powers. This is where we see that the basic validity rule cannot fit within Searle’s theory as it stands. Finally, I locate the difficulty within Searle’s account of codification, dismiss a possible reply, and suggest an emendation.

1. Codification

Searle tells us that institutional facts are a subclass of social facts, which means that they depend upon collective intent (1995, 26; 2010, 156). Some of these social facts are about practices that are “codified,” which lends them the properties of an institution, although uncodified, “informal” institutions are also possible (Searle 1995, 53). Codification is therefore Searle’s answer to the problem of validity: institutional facts get their official status via codification.

Searle’s notion of codification is more general than, but likely based upon, the more narrow legal sense. He does not give a more precise explication, other than to note that codification involves some kind of formal declaration. In law, “codification” is usually understood to refer only to the legislation of a legal norm, or perhaps the administrative enactment of a regulation. But it is clear from Searle’s usage that his notion would apply equally well to legal norms imposed by judicial action in that they are similarly formalized. I will use the term in Searle’s broad institutional sense even when talking about laws or legal systems, although we should keep in mind that legal codification is one species of this broader notion.

For Searle, as for others, observer-relative social facts are distinguished from observer-independent brute facts (Searle 1995, 27). The former are facts whose truth is determined by our social environment and hence dependent upon our behavior or recognition in some way (although this does not threaten their epistemic objectivity, Searle 1995, 8-12); the latter are facts about the world whose truth is independent of our understanding. That an object is made out of wood and metal is a brute fact; that it is a hammer is a social fact.

11 “[A] social fact is simply any case of collective intentionality involving two or more animals” (Smith & Searle, 2003, 304).
In Searle’s theory, an institutional fact is created by the collectively intended assignment of a status function (usually, “X counts as Y in context C”\textsuperscript{14}) to something (usually conceived as a practice, but it can be just about any social phenomenon) via a constitutive rule (1995, 23-26, 40-45).\textsuperscript{15} The status function conveys deontic powers, which “provide desire-independent reasons for action” (Searle 2010, 23). In other words, an institutional fact reflects a certain kind of social activity that can take place over time by creating, identifying, or changing rights and responsibilities among the participants. It is created when we have a set of rules that constitutes the activity by imposing a status upon the behavior or related objects or events.

While Searle claims constitutive rules create the possibility for the activity and thereby differ from rules that simply regulate the activity (1995, 27),\textsuperscript{16} Joseph Raz has persuasively argued that all rules are both constitutive and regulative in Searle’s sense, depending on the descriptions one gives of the behaviors performed in conforming to the rule (Raz 1975, 108-11). Any distinction that can be maintained would therefore attach to the functions that the rules perform, the descriptions we give of them, and the behaviors we perform in following them. Using this distinction therefore doesn’t suggest that there are two different types of rules; it only calls attention to the way a given rule is operating in a specific context. While Raz sees this as a good reason for preferring the notion of “power-conferring rules” (1975, 111, following Hart), we will stick with Searle’s terminology for simplicity of analysis, keeping in mind that it should refer to a description of a rule’s use, rather than a distinct type of rule.\textsuperscript{17}

While only facts about objects, practices, or events assigned functional statuses they could not otherwise have had are candidates for institutional facts (Searle 1995, 88-89), that imposition of function also must be collectively intended (in a specifically empowered context).\textsuperscript{18} Otherwise I could create a new institutional fact simply by looking at my pen and declaring it to be the scepter of power to be used in all official state ceremonies.\textsuperscript{19}

\textsuperscript{14} (1995, 28). Conformity with this formula is not absolutely necessary. It serves as a “mnemonic to remind us that institutional facts” require a collectively intended assignment of a special status to the object (Smith & Searle, 2003, 301). See also Searle (2010, 19-20).

\textsuperscript{15} See also Smith and Searle (2003, 302-03), discussing the reasons for not limiting the recipients of status functions to objects.

\textsuperscript{16} Both Searle and Hart understood that constitutive rules can also regulate the activity. See Searle (1969, 33-34), cited by Marmor (2006a, 350, interpreting Hart).

\textsuperscript{17} I thank an anonymous reviewer for this journal for alerting me to the need for this clarification.

\textsuperscript{18} We collectively intend for the practice or object to take on a special functional status. Our intention is collective in that we collectively impose the function on the object and we intend (or recognize, Searle, 2010, 56-57) that object to fulfill the function we impose upon it, (Searle, 1995, 23-26; 2010, 59-60). For criticism of Searle’s reliance on collective intentionality as the backbone of his theory of institutions see Turner (1999, 223-29). I will go into greater detail about Searle’s notion of collective intentionality below. For now, it is enough to understand that his notion of collective intentionality is one in which each individual possesses an intention that is identical to those of the others in the group, and that the intention is irreducibly collective in that its content references the group (Searle, 1995, 23-26; see also 2010, 47-48). It should also be noted that Searle avoids any metaphysically mysterious, Hegelian-style notions of collectivity here. The intentions are held by individuals; their content is collective (1995, 25). See also Zaibert (2003, 62-63, noting the inadequacy of Searle’s treatment).

\textsuperscript{19} In this, institutional facts as understood by Searle differ from works of art and many other kinds of artifacts that we would not consider to be institutionally defined. In many of those other cases, the individual intention of a single creator may be enough for status facts about the object to be made true.
The imposition or assignment of function\(^\text{20}\) is simply the use of an object to fulfill a purpose.\(^\text{21}\) An important subset of these assigned functions is symbolic, in which the ascribed function is to “stand for” or “represent” something else (Searle 1995, 21).\(^\text{22}\) Furthermore, these assignments can be unconscious, even while they must be intentional (and hence at least capable of being understood) (Searle 1995, 21-22). Since it requires collective intentionality, we can call the creation of an institutional fact a collective imposition of function.

The imposition of status functions takes two forms. In one form, collective intention attributes the function to the recipients individually. This is true for cocktail parties (Searle’s example). Generally, we are only at a cocktail party when most of the people at the gathering consider it to be a cocktail party. It is not that they must think, “This is a cocktail party.” Rather, they must treat the activity in which they are engaging as a cocktail party and thereby collectively attribute the cocktail party function to the activity.\(^\text{23}\) In the other form, our collective intention attributes the function to the recipients as a type. This is true for money (again, Searle’s example). A given piece of paper or metal counts as money because it fits a certain description and the status of money has been conferred upon anything that fits that description (although this need not be conferred by law). So, in the former case the ‘X’ term in the “X counts as Y in context C” formula is a particular instance, while in the latter it is a description that can fit multiple possible tokens. In the latter case, then, there must be something that fixes the type and determines how to recognize tokens as instances of the type, bearing its status function. This is done through codification.

Codification is therefore a formalized process of declaration for attaching or creating statuses. (This is the closest we can get to a definition based on Searle’s treatment. While I acknowledge that calling the process “formalized” may be almost as opaque as codification itself, the idea is that codification happens by following some established rules or procedures for making the status ascription.) Adding codification moves us from token-status ascriptions for informal institutions and social facts more generally, to type-status ascriptions for formal institutional facts.\(^\text{24}\) This means that, where codification is present, the assignment of function in the constitutive rule “X counts as Y in context C” itself becomes institutionalized. Each cocktail party must be collectively treated as a cocktail

\(^{20}\) Searle uses “imposition” and “assignment” of function interchangeably (1995, 13).

\(^{21}\) Ibid., distinguishing “imposed,” “agentive” functions from “discovered,” “non-agentive” functions (also 2010, 58-59). See also Fotion (2000, 179-80). In Searle’s words, it is the “ascription of . . . the use to which we intentionally put these objects” (1995, 20, emphasis in original).

\(^{22}\) These are Searle’s use of the terms.

\(^{23}\) While Searle clearly states, “Part of being a cocktail party is being thought to be a cocktail party . . .” (1995, 88-89), I believe a charitable understanding of his theory is that there must be a self-conscious attribution of the function of a cocktail party to the event, not that anyone must think of the gathering as a cocktail party in so many words. I will return to this point below.

\(^{24}\) More precisely, since the ascription of function defines the social or institutional entity, there is a self-referentiality to that definitional process that would otherwise be circular. (To be a cocktail party is to be treated as a cocktail party.) This process is saved from circularity because we bestow the name upon the practice when we name the function, which is itself simply a “placeholder for the linguistic articulation of all [the] practices” that comprise the function being ascribed. Whereas generally this self-referentiality applies to each token of the practice (in that we ascribe the status function to each token individually), codification allows us to ascribe the status function to practices by types. Ibid.
party in order to be a cocktail party, but a certain twenty dollar bill can still be a piece of currency even if no one ever comes into contact with it.25

Searle states that the test for the presence of an institutional fact is whether the constitutive rules imposing the status function could be codified explicitly (1995, 87-88). Where that codification is possible but has not yet been done, the institution is merely informal (Searle 2010, 91; 1995, 88).26 Among other things, an informal institution lacks clear criteria for identifying invalid instances or applications of the institution’s norms. (There are not really any rules for what counts as a ‘real’ cocktail party and what does not.) Hence the use of codification to distinguish between formal and informal institutions is precisely what allows for the presence of criteria of validity within the institution.

While Searle remains vague about how he conceives of codification itself, claiming only that to get type-level functional ascription, codification must be “official” (employing the quotation marks: 1995, 53) and explicit,27 he clearly envisions legal enactments as paradigm cases. We can put these notions together to suggest a more complete picture: there will be a person or body empowered (by a constitutive rule of the institution) to establish the assignment of function (within a given domain of discourse) as at least one of its tasks.28 Furthermore, that person or body must issue a symbolic representation of the codification, usually by reducing it to words. Of course, we will need other institutional facts below these to pick out what is “official” (which gives rise to the problem addressed in this paper).

We cannot expect clear necessary and sufficient conditions for Searlean codification as there is no “sharp dividing line between social facts in general and the special subclass of institutional facts” (Searle 1995, 88). However, more recently Searle has focused on the use of a specific kind of speech-act, a “Declaration,” to create the institution.29 Hence, we can extend the notion of codification to include the need for an official declaration of some kind, while uncodified, informal institutions can form without the need for declarations.30

Practices whose status functions remain uncodified have the advantages of “flexibility, spontaneity, and informality,” which are lost through codification (Searle 1995, 88).31

25 “Codification specifies the features a token must have in order to be an instance of the type [where we ascribe the status function to types rather than tokens].” Ibid.
26 In conversation, Searle confirmed that he thought the difference between institutions and mere social facts turns upon whether codification of the status function is possible; where it is possible but has not yet taken place, the institution is informal.
27 Contrast this with the notion of codification employed by David Lewis (1969, 104, explaining that the rules of a game are codified when written down and such codifications can be used to teach the game).
28 This empowerment itself will usually be the result of an earlier codification. Additionally, the person could be artificial and the body could be the entire community.
29 Searle (2010, 85-86, capitalizing “Declaration” to indicate that it is a characteristic kind of speech act). This focus is meant to help explain the possibility of institutions that are not mapped onto pre-existing entities (so called “free standing Y-terms” in which there is no X in the “X counts as Y in context C” formula, 2008, 454; Smith 2003, 19-25).
30 In some cases, there can even be institutional facts without institutions. Searle’s example of this is a line of stones (the remains of an ancient wall) that continues to be treated as a boundary (2010, 94-95).
31 Andrei Marmor has distinguished between two kinds of codification: “Legislative codification of rules purports to determine, authoritatively, what the rules are. In contrast, encyclopedic codification only purports to report what the rules are, without actually determining their content for the future” (2007, 607). The phrase quoted in the text and other passages of Searle’s (see e.g., infra, n. 33) make it clear that his notion of codification corresponds most closely to Marmor’s legislative form.
Hence codification involves some notion of fixity that entails the loss of these characteristics.

We can get some insight into this aspect of codification by comparing Searle’s analysis to Hart’s notion of the difference between a “primitive” or “customary” legal system and a “modern” legal system. The primitive system consists only of what Hart calls “primary” rules, which correspond to what Searle would call merely regulative rules (of course, the wider, non-legal context will have many constitutive rules and keeping in mind that we have rejected any sharp distinction here). The modern system includes “secondary” rules to constitute the legal institutions and offices, providing for official codification. The customary system suffers from the defects of “uncertainty” about the precise meaning or application of the rules, the “static” nature of the rules in a system that does not have any institutional procedures for their intentional change, and the “inefficiency” of having no institutions to enforce them (Hart 1994, 92-93). While they are not dealing with the precisely the same subject matter (legal systems for Hart, formal versus informal institutions for Searle), the mark of the development of the “modern” legal system is precisely a set of “secondary” rules that provide for official codification.

While Searle notes that there is no bright line distinction between social facts generally and institutional facts, he also claims, “The characteristic institutional move, however, is that form of collective intentionality that constitutes the acceptance, recognition, etc., of one phenomenon as a phenomenon of a higher sort by imposing a collective status and a corresponding function upon it” (1995, 88). The example of cocktail parties might bring this process further into the light. Initially, cocktail parties were simple social facts. People came together to drink alcohol and engage in witty banter, without thinking anything special of it or attaching to it any special statuses, rights, or duties. They did so intentionally in the sense that they intended to drink alcohol and engage in witty banter at a gathering of people, and that required a similar intention on the part of others. But they did not necessarily consider it a special event to do so. Many, perhaps even most, cocktail parties are still like this, especially those that seem more to be cocktail parties in retrospect than in the planning. Other cocktail parties are special events in the sense that hosts decide to throw a cocktail party: they send out invitations in which they request replies. People who receive the invitations know that they are invited to a cocktail party, and that it is incumbent upon them to reply. If they attend, they do so knowing what to expect when they get there: that the hosts will provide a certain atmosphere and some alcohol. These special events have risen to the level of informal institutions. At this point, there is much more self-consciousness about the cocktail party. It does not rise to the informal institutional level of the special event unless people consider it as such and engage in “the acceptance, recognition, etc., of one phenomenon as a phenomenon of a higher sort by imposing a collective status and a corresponding function upon it.” Searle also invites us to imagine what it would take for cocktail parties to rise to the level of a formal institution (1995, 88). Some official body would have to codify rules specifying which gatherings are to be considered cocktail parties, and what the rights and

---

32 Their apparent differences over the advantages and disadvantages of a fully developed institutional system can be harmonized somewhat by looking at the perspectives from which purposive enactment might be a strength or a weakness. A fully developed legal system is less static than a system of entrenched customs because it provides for purposive change. But it is also less flexible than a system relying on token recognition of statuses because it requires following an institutionalized procedure for change.
responsibilities are of hosts and invitees. The mark of what makes it pass from an informal to a formal institution is thus whether its rules are codified such that we can say which instances are valid instances of the kind and which (if any) are merely counterfeits.

2. The Basic Validity Rule and Collective Intention

The problem arises when we get to the most basic constitutive rule that identifies which codifications are valid. Codification in complex institutions is iterative (Searle, 1995, 80f, 116). The only reason that a piece of currency is money is that there is a codification of the appropriate assignment of function. The only reason that codification is itself legally effective is the codification of another institutional fact lying behind or underneath it, something of the general form “statements written in these books (or spoken by these people) count as legally binding when passed according to the rules listed in this foundational document.” On these points Searle and most legal positivists would agree. Notions of legal validity are similarly dependent on more basic constitutive rules, which in turn are dependent on still more basic rules (Hart 1994, 107). But Searle’s framework does not give us an adequate way to account for the most basic validity rule.

The most basic rule directing legal officials how to recognize valid laws cannot itself be recorded as a law within the legal system it validates (Shapiro 2011, 86). If we were to try to legislate or legally codify the basic validity rule, we would need to have another – even more basic – rule telling us to recognize the legal validity of the rule we just tried to legislate. And that rule would then take its place as the basic validity rule. Furthermore, since it is an ultimate rule, it is not dependent on others for validity; questions about its legal validity are misplaced since it is constitutive of legal validity (Hart 1994, 109; Shapiro 2011, 90; Coleman 2011, 8 n.7; Coyle 2006, 420). It cannot be formulated

33 While it is possible to envision informal institutions, once formal codification takes place, specifying the characteristics a token must have to be a member of the type, informal institutional examples are no longer possible for that entity (or are at best counterfeit instances of it), although they still might exist as simple social facts. Searle’s example of the Korean War shows the unavailability of informal institutions when codification is present (1995, 89). American officials were very anxious that it be called a “UN Police Action,” so as to avoid the risk that the military action would be found unconstitutional. It was still a war in the basic social fact sense (and so it is not incorrect to call it a war), but it was not a war in the institutional sense. Given this analysis of Searle’s, we cannot say that the Korean War was an informal institutional fact. Its status as a war was studiously and purposely not made an institutional fact, and was purposely left simply a social fact. (It might still have been an instance of a different kind of institution as well: a “Police Action.”) While the example of the Korean War invites us to imagine codification in terms of its legal paradigm of legislation, we should remember that Searle’s notion is much broader than legal enactment. We can also see examples of this sort in sports and religious rituals, where a token performance can fail to be a member of the ‘officially recognized’ type because of a flaw in its pedigree or execution. Hence codification, in specifying what features a token must have to be a valid member of the type, sets forth standards of validity that would exclude “informal” exemplars as counterfeits (Hindriks 2003, 203-05).

34 This is not meant to be a quotation from a legal instrument. The basic validity rules are generally understood to be customary (or practice-theory) rules. They are not generally thought to be reducible to a legal formula and recorded as law since they would require a further more basic rule to enable that recording to serve as legally valid (and hence there would need to be another more basic rule below). See Hart (1994, 111), Alexander and Schauer (2009), and as discussed immediately below. See also e.g., Coleman (2011, 8 n.7), Coyle (2006, 420), both citing Hart (1994, 109, claiming that questions of validity cannot arise about the rule of recognition itself).

35 See also e.g., Greenawalt (1987, 622), reprinted in Adler and Himma (2009, 2).
as a law within the legal system precisely because it is the rule that determines what counts as law in that system. Its existence is a matter of fact inherent in the practices of officials (and sometimes citizens as well), which reflect their acceptance of it. According to Hart, the rule of recognition is not even usually stated explicitly (1994, 107).

Clearly, this means that the basic validity rule cannot be codified in the legal sense. But it also raises a problem in Searle’s broader sense of codification. The way the basic validity rule sits outside the set of laws it validates means it does not fit into Searle’s analysis: it is the constitutive rule of official (Searlean) codification for legal institutions, specifying which legal enactments “count as” officially codified in that jurisdiction.

The behavior of legal officials with respect to their determinations of legal validity is both rule-described and rule-governed (or guided) (Searle 1995, 139; Shapiro 1998). The content of the rule is a description of criteria legal officials use to determine which laws are valid within their system. But it also serves as a reason for those officials to use those criteria where they are either motivated or have good reason to do as other officials are doing. Their authority is a mutually supporting network of recognition under the basic validity rule.

The idea of accepting a description of the behavior of officials as a reason for them to behave in accordance with the criteria exemplified by that behavior reflects the characteristic transformation of a social fact into an institutional one under Searle. The description of their behavior now has a new status that includes the creation of deontic powers, at least from the institutional point of view (Searle 1995, 100-01; 2010, 92, 105-06, 118, 167): officials are institutionally obligated to act in conformity with the basic validity rule. (Whether this institutional obligation is a moral one depends upon the wider moral

36 Scott Shapiro notes that there is a sense in which the rule of recognition is a rule of the system it validates, but it is not a law of the system it validates. “[T]he law of a particular system just is the set of rules that officials of a certain system are under a duty to apply and the rule of recognition sets out the content of this duty” (Shapiro 2009, 240; 2011, 85 (this passage is emphasized in the texts), and citing Raz 1979, 92).

37 Hart (1994, 101-02, 107-10, 256). See also Dickson (2007, 11). As I discuss below, this statement is not intended to endorse Hart’s practice theory explanation of the basic validity rule.

38 One might think that the U.S. Constitution specifies the content of the basic validity rule for United States law, and that the Constitution was ratified according to an official process and hence codified in Searle’s sense. There are several problems with this objection. For one, following Hart’s analysis, it is more correct to say that the rule of recognition for the U.S. is something like, ‘whatever rules are adopted and applied according to, and are consistent with, the rules set forth in the U.S. Constitution’ (see 1994, 107.) That is, since the Constitution is the “supreme law of the land,” there must be an additional rule (which is itself not a law) that officials are following to treat the Constitution as the supreme law. Shapiro articulates another problem with seeing the Constitution as setting out the content of the rule of recognition: he claims that the rule of recognition is a duty-imposing (although still secondary) rule, while the Constitution is mostly power-conferring (2011, 85-86). See also generally Greenawalt (1987), reprinted in Adler and Himma (2009).

39 We will see below that this is why Searle cannot accommodate the basic validity rule in his “Background” capacities (1995, 145).

40 Shapiro criticizes Hart’s “practice theory” description of the basic validity rule as a category mistake, claiming rules are abstract entities, while practices are spatially and temporally located (Shapiro, 2011, 103). This argument has received its share of criticism (e.g., Sciaraffa 2011, 610; Gardner and Macklem 2011; Kramer 2013, 380-83; see also Green 2013; offering a qualified defense). I sidestep this issue here by focusing on the basic validity rule’s normative nature without taking a position on from where that normativity is properly derived. I address this issue more directly and offer my own reply to Shapiro on this issue in Ehrenberg (2018).
legitimacy of the institution). It is this characteristic that requires us to see the basic validity rule as more than a simple social fact. Yet it does not fit within Searle’s framework to describe its institutional nature, neither as an informal institutional fact, nor as a formal one. It would be a requirement within Searle’s theory for the basic validity rule to be codified, and yet it is impossible to do so because of its special characteristics.41

The reliance of legal positivists upon the acceptance of the basic validity rule as the cornerstone of a legal system mirrors Searle’s analysis that legal systems depend on collective acceptance rather than force (2010, 163, 165).42 One might think that since the rule itself depends for its existence only on the acceptance of certain key officials, it does not depend upon a broader collective intentionality. This would seem to run afoul of Searle’s claim that the collective who must accept the institution consists of those who are members of or subject to the institution. The idea might be that since the basic validity rule only requires a relatively small number of key officials to follow it, it is not an object of society-wide collective intentionality. In the case of the institution of law, everyone in society is subject to the institution and yet only the key officials need accept the basic validity rule.

This is not the case for two reasons: First of all, the rest of the populace tend to defer to those officials in the determination of legal validity. This is analogous to a linguistic or epistemic division of labor. We rely on scientific experts to determine the atomic structure of gold, allowing us to rely on their expertise (or that of others familiar with the scientific markers) to decide for us which references to or exemplars of gold are correct and which are really just iron pyrite.43 Similarly, we rely on legal experts (officials) to determine which putative legal rules are valid under our legal system.44

Secondly (and this bears some emphasis), it is not necessary for the collective that possesses the relevant intention to be identical with the collective to which the institution is meant to apply. If there is a splinter group off in the woods claiming that the government is illegitimate because of some perceived inconsistency between the details of its formation and the procedures found in foundational texts, we do not immediately say that the government’s legitimacy is a subjective matter. That the key officials accept the

41 While related, the problem we are focused on in this paper is not the same as the chicken-egg problem discussed by Shapiro (2011, 40-41). That is the problem of who authorized the rule makers to make the rule specifying who is authorized to make rules (although put in terms of legal norms). Hart’s rule of recognition can be seen as a solution to that problem, saying essentially that a customary (practice-theory) rule provides the mutually supporting network of recognition necessary to establish the authorization to make and interpret legal rules (although Shapiro attacks Hart for offering this solution – see previous footnote for details). While that is a general problem of authority in formal contexts, our problem is one in the ontology of social institutions: how an informal, customary rule can provide the fixity needed for the validity standards required in formal institutions.

42 Searle offers the Los Angeles riots of 1992 and the response of police, who were unable to exercise authority once the number of law breakers grew too large, as evidence of this point (1995, 90-91). This is almost identical to Hart’s point that for legal force to be applied against anyone within a jurisdiction, there must already be widespread acceptance of the legal authority (1994, 201). See also Shapiro (2009, 251), detailing the need for trust in legal officials for a legal system to operate.

43 I purposely avoid here the question of whether the reference is correct only because the experts (of the linguistic community) say so, or if they say so only because it is correct. The point is that the linguistic community relies on experts in either case.

44 Admittedly, in the case of scientific experts, their expertise is entirely epistemic, while for legal officials some or much of their expertise is constitutive of the content of law. (They may be ‘experts’ only in the sense that the institution assigns this role to them, or because they had a role in the formation of the law.) But this distinction does not threaten the analogy made here.
legitimacy of the government is enough to make it so (by its own lights) since the institution itself gives them the power to make the determination, and that institution is supported by at least the tacit acceptance of the populace. This is not to say that the splinter group does not have a valid point, only that the legal legitimacy of the legal system is not a subjective matter simply because some people disagree.

This is similar to the decision of a referee in certain sporting events. Her determination about a disputed play suffices for the official record of the game, regardless of what players or spectators might think. This is not to say that the referee is infallible. Players or spectators who are familiar with the rules might dispute her determination on the basis of their own perception and interpretation of events. But they cannot reasonably dispute that the referee is given the power to make the final determination of the play’s official status within the game. Of course, the referee is just one individual, so there must be a network of collective intentionality that puts her in that position and gives her the power to make the official determination.

The basic validity rule is necessarily an object of collective intentionality. Key legal officials must operate with the same rule in order to have a coherent legal system; they follow the rule because it is the rule that the group accepts (although they may also have other reasons upon which they rely, and they need not be fully psychologically self-aware in their use of the rule). The motivation for the officials’ acceptance might be personal, but the content of the intention they have in doing so is collective: ‘We intend that rules that meet these criteria will be our law’ (although, again, this is not necessarily consciously explicit).

While there are several different understandings of collective intentionality in the literature, I do not believe much turns here on which kind of collective intentionality is correct, so long as it allows for the kind of implicit or un-self-conscious sharing of intentionality that both legal positivists and Searle see as prevalent in legal institutions. Searle’s notion of “we-intentionality,” in which the content of each individual’s intention is identical to that of each other individual (1995, 24-26; 2010, 47-48), has the advantage of streamlining coherence among the officials about the validity rule since it would then be possible for their intentions regarding it to be identical.

Another advantage is that it places the interdependence and group reliance of the basic validity rule in sharp relief, highlighting the need for coherence. While it would be possible for a given official to be wrong about the content of the intention, the fact that it is

---

45 See e.g., Himma (2001, 289), noting that the rule of recognition does not bind ordinary citizens since it is not aimed at them, and Shapiro (2011, 85).
46 But see Dworkin (1986, 122), and Adler (2006, 745-47). While these views raise doubts about the shared content of the basic validity rule, they do not necessarily threaten its status as an object of collective intention. See Shapiro (2009, 260-61), showing that these views do not threaten the reliance of the rule of recognition on shared plans.
47 Contra Dickson (2007, 22-23), who argues that the basic validity rule cannot be a Lewisian convention (partially) because those who follow the rule do not generally do so simply because others do so (thereby disqualifying it from being a convention in Lewis’ sense), see Postema (1982, 177-78), arguing that one may have his own, nonconventional, reasons for accepting a Lewisian convention. See also Marmor (2001, 199-203).
48 Searle does note that unconscious intentional states must, at least in principle, be accessible to consciousness (1995, 7).
49 Criticism of notions of collective intentionality that do not reduce it to individual intentionality can be found in Bratman (1999, 93-161).
even possible to be wrong supports the notion that the intention held by the vast majority of the officials is identical or nearly so. While some disagreements among officials as to what counts as valid law are inevitable, the nature of the legal system itself requires that these disagreements be resolved by treating disfavored decisions as mistaken applications of the validity criteria. Each official is tracking the validity practices of the others with the understanding that incorrect applications of the criteria represented by those practices will be rejected as invalid. Hence the collective intent is to have a uniform set of practices that is then viewed as normative by those officials.

3. Problems in Formalizing the Assignment of Function

The key move in codification for Searle is the formalization of the assignment of function. What makes a social fact an institutional fact is that the constitutive rule assigning the status function could be codified “explicitly” (Searle 1995, 88). (Again, where the status function is not yet codified, but could be, the institution is only informal.) For Searle explicitness is a necessary component of codification. As mentioned above, codification removes the flexibility, informality and spontaneity of social practices by formally recording the constitutive rule.

The characteristics associated with uncodified social facts are ones we neither want nor expect to see in the foundational norm that our officials use to decide what counts as valid law. We want and expect our basic validity rule to be relatively stable and definite. Cocktail parties can arise informally and spontaneously. If one or two or even more of the attendees happen not to think of it that way or think that they are all at a different kind of social gathering, it does not much matter. But if many key legal officials have very different ideas about what makes any putative law a valid legal requirement in their jurisdiction, then the legal institutions themselves are thrown into doubt (Hart 1994, 122, 148).

Of course, there may be some disagreement over interpretation of the basic rule or of intermediary facts upon which the validity of other laws are based (such as the meaning of constitutional terms). But if the assignment of function itself is not relatively rigid, then we will have a much more radical difference of opinion among officials as to what counts as law. We might not be able to tell when we are married, or divorced, or contractually obligated because we would not be able to tell which explicitly legally codified enactments are officially valid.

The law would not then be able to fulfill its Searlean codification purpose in formalizing institutional facts, nor would it be very good at guiding behavior. If a lack of codification of the basic validity rule means that the law cannot fulfill any of its main social functions, then it would seem that Searlean codification of the basic rule is a necessary step in the development of a legal system.

---

50 Of course, where “mistaken” decisions come to be accepted we might say that the accepted validity criteria have been changed.

51 Shapiro’s planning theory of law highlights the importance of this caveat (2009, 265-66, showing how accounting for differences in interpretive methodology allows for disagreements regarding the criteria of legal validity). See also above n.9 and ibid. (noting that a portion of the rule of recognition will still be ultimate).

52 For an antipositivist like Dworkin (who does not believe in an absolute rule of validity), this distinction may be one more of degree than of kind. (1986, 91). See also Marmor (2001, 197-98).

53 Remember that this is not meant to suggest a legal recording, which is conceptually impossible since the basic validity rule cannot be understood to be legally valid or invalid.
This is the surprising thing about the basic validity rule when viewed through a Searlean lens: “Its informality does not make the entire legal system as evanescent as a cocktail party.”54 The explicit laws of the legal system are codified (in both senses), relatively stable and definite. But they are based on an uncodifiable informality. They are legally valid (and therefore “officially codified,” with a successful assignment of their status function) only if they conform to that informal basic validity rule. Under Searle’s analysis, the fact that the basic validity rule is not codified should make the entire legal system “as evanescent as a cocktail party,” but it does not.

To be clear, the content of the basic validity rule is undoubtedly an institutional fact. It is still an object of collective intent and sets the criteria for the correct or successful assignment of function, imposing the status of law on those legal enactments within its ambit. It is still a description of social practices counted by officials as a reason to conform to those practices. It makes the characteristic institutional move of creating deontic powers by specifying the validating roles of certain key officials. The problem is that, for it to do its job properly, it both must be and cannot be codified.

Recall that codification is the way to get an institutional fact to move from token-level status ascription to type-level status ascription. Without codification, each basic validity rule must be considered to be such a rule in order to be one.

This might not look like much of a problem. After all, basic validity rules are probably more like cocktail parties. We consider each token basic validity rule in its capacity to fulfill its functional role of setting the validity standards for the legal system it validates. Each legal system has its own basic validity rule and the officials of each of those systems must consider it as a token in its role. Officials within each system consider a given set of facts (e.g., that a piece of legislation was passed by the relevant majority and signed by the executive) to set the validity conditions for their system. So those officials are treating the basic validity rule at the token-level as conferring legal status on the enactments it calls upon them to identify as such.

However, Hart and other theorists explicitly acknowledge that these rules are not something that officials are psychologically self-aware about when they accept them. They do not think to themselves, “I am following a basic rule that tells me to look to the U.S. Constitution and decisions of the Supreme Court to make validity determinations,” although they are self-conscious about applying the rule (Hart 1994, 107-08, noting that the rule of recognition is an external fact about the legal system).

For clarity and ease of discussion, let us make a distinction here between a “self-conscious” endorsement and application of a rule, and one that is “self-aware.” (I am using these terms in a special way here and don’t intend any implication that these are reflective of usage in psychological literature.) An official’s determination of validity in accord with the basic validity rule is “self-conscious” in the sense that she knows that she is applying rules to which she has access and thereby making a determination of validity. But it is not generally “self-aware” in the sense that she is not usually psychologically aware she is following an unwritten basic validity rule. She perhaps thinks instead she is simply following the foundational document or practice. Another way of putting this would be to say that one self-consciously applies a rule when one intentionally performs the actions (or forms the beliefs) that the rule directs, while one is self-aware when one has psychological access to the semantic content of the rule. That one can apply a rule

54 This way of putting the point is thanks to a blind reviewer.
without being fully aware of its content should not be too surprising, as many instances of etiquette are likely of this kind: an awareness that the behavior is demanded in that situation, but rarely aware of the exact content of the rule one is following.

The distinction I make here could be seen as a simplification of a pair of more complicated ones advanced by Gerald Postema. He makes two distinctions that cut across one another. The first is the distinction between “personal” and “logical” perspectives or points of view. The personal is understood as “psychological or epistemic ‘sets’” belonging to actual or imagined people. The logical is a way of describing “regions of logical space.” He refers to these as “distinctive frameworks for inquiry.” The second is the distinction between the “theoretical” and “practical” domains of inquiry, the first being occupied with discovering what is the case, the latter occupied by what to do about it (Postema 1998, 342). The distinction I draw is meant to capture the difference between the theoretical and practical domains of inquiry within the personal perspectives of the officials. Since officials are occupied with the practical domain of applying the basic validity rule, they generally have no need to attend to the theoretical descriptions of those practices.

Officials are unlikely to think of themselves as discovering the criteria of legal validity in an unwritten rule unless they happen to have read Hart or other legal philosophers. Rather, they are likely to think that the foundational document or practice is the most basic rule. They use the basic validity rule to make explicit and conscious determinations of legal validity; they are just not usually aware they are using an unwritten basic rule to do so.

4. An Insufficiently Explicit Intention

We expect some explicit collective intention to codify the basic rule in order to lend it its institutionality. But the possibility of this for the basic validity rule is undermined by Searle and legal positivists’ point of agreement that institutions are based on social facts. For Searle, any social object “is just the continuous possibility of the activity” (1995, 36) that lies behind the object. Institutions, as kinds of social objects, do this by articulating the normative contours of what actions meet the criteria of membership to be counted as exemplars of that practice or activity. This echoes the notion shared by some positivists that the basic validity rules are really just a description of shared practices among the legal officials, a description of what the group of officials takes to be valid that they internally adopt as reasons for conforming their validity decisions to those descriptions.

---

55 This qualification explains why I say that officials do not generally have access to a formulation of the basic validity rule even as they follow and implement it (although it has recently become apparent to me that this may not be true in the UK, where judges routinely cite Hart in their opinions). It might seem that the problem raised in this paper would be moot if legal officials were all legal positivists. But the problem is conceptual in that we have legal systems in which officials are not knowledgeable about positivism and yet they still function as officials (implying that such a situation is possible). Both Hart and Searle purport to explain those systems and yet officials are not generally self-aware of the basic validity rule. Furthermore, even if it were the case that some officials in every existing legal system are self-aware (itself a great stretch to the imagination), the institutionality of the system depends upon many officials who are not self-aware.

56 See, e.g., Hart (1994, 107-09). See also, Marmor (2006b, 3-4), emphasizing the conventionality of the basic validity rule as a way of explaining its institutionality.
For Searle’s analysis of institutional facts to hold of the basic validity rule, the officials would have to intend collectively that something like “passed in conformity with our jurisdiction’s officially recognized foundational document and/or practices” (appropriately cashed out with more specifics) constitutes the basic validity rule for their jurisdiction. But if this formula is in reality simply a description of a practice and not the psychologically explicit object of anyone’s intention (although its application is intentional), then it is difficult to see how institutionalization of the basic validity rule is possible.

Recall Searle’s claim that for a social fact to be institutional (at least informally), the constitutive rule must be capable of explicit codification. So we might consider it enough that the above formula could be made explicit, even if not legally codified.

But then a more serious problem arises. We have seen that the formula is not usually endorsed as such in a psychologically self-aware manner, but is rather a description that theorists impart to the legal system as the practice of its officials. These theorists are, of course, explaining the system from an external point of view. The content of the basic validity rule is not generally psychologically available to the officials, even if it is a perfectly accurate description of what those officials are doing. Hence this particular constitutive rule is not capable of explicit codification since those that do the codifying (the officials) do not generally have access to any formulation of the rule.

To summarize, there are three possible ways to understand the basic validity rule from within Searle’s theory: 1) as a full-blooded institutional fact, 2) as an informal institutional fact, and 3) as a simple social fact. Yet none of them seem open to us at the moment.

(1) The basic rule cannot be a full-blooded institutional fact because there is no collectively intended assignment of the status function. The “X” term in the “X counts as Y in C” formula would be a description of the criteria legal officials use in making determinations of legal validity. The “Y” term would be the conversion of this description into a reason-giving behavior-guiding rule. The “C” term would be the legal jurisdiction in which those particular officials operate. (That context itself is a creation of the legal system and hence an institutional fact.) The problem is that we have no constitutive rule assigning the status function to the description of the officials’ behavior. We have good reasons to think that the basic validity rule is an institutional fact; we just cannot use Searle’s theory to explain it as such.

(2) The basic validity rule cannot be an “informal” institutional fact because we could not codify the constitutive rule even if we could locate one, and the availability of a rule that could be codified is a necessary characteristic of an informal institution.

(3) If, instead, we wish to say that the validity rule is a simple social fact, we run back into the problem that we cannot then understand how a simple description of officials’ practices is creating reasons and guiding their behavior, since a simple uninstitutionalized social fact does not carry any deontic powers. The officials are certainly imposing a function on the description of each other’s validity practices: they are treating those practices as reason-giving. So the basic validity rule does not fit any of Searle’s categories.
We would not usually be bothered by the fact that officials are not psychologically self-aware when following a basic validity rule. Judges believe themselves bound by their constitution to apply the relevant sources of law, not that they are following a more basic rule that tells them to look to their constitution to find the relevant sources. Much theorizing about the world around us carries with it the implicit claim that people are engaging in behaviors or holding beliefs without fully realizing their theoretical characteristics. In Searle’s use of collective intentionality, however, we have a claim that (at least in some cases) people must be aware of the status they collectively intend in their assignment of function. Cocktail parties are not cocktail parties if we do not collectively intend them to be cocktail parties. Pieces of paper cannot function as money if there aren’t at least some officials who collectively intend for them to be used as money. To give rise to an institutional fact (even in an informal institution), we have to be sufficiently self-aware of our ascription of functionality to the practices in which we are engaged. We must at least be aware of treating the gathering as a cocktail party by self-consciously behaving as we do at cocktail parties.

The problem we have been examining does not arise because the officials fail to think consciously of the basic validity rule as a basic validity rule. Rather, it arises because the officials do not generally have sufficient intentionality to treat the description of their validity practices as a rule for determining validity. To be sure, they follow the validity rule because other officials are doing so (if they have internalized it as reason-giving), but they are not psychologically aware of giving it the status of a validity rule. For them it continues to be a practice that they follow because others do so; there is no recognition of the practice as counting as something else.

Searle would be the first to remind us of the linguistic and institutional division of labor. One might therefore think that the highest ranking legal officials, who pass on issues of validity, are relying on the expertise of philosophers in order to intend successfully and collectively that their basic norms fix the criteria of legal validity for their legal system. This cannot be a satisfactory answer. First of all, when we rely on the scientists to do the background scientific work for us to speak more accurately about the world, those scientists actually have to endorse the scientific propositions that lie behind our statements. When philosophers talk about legal officials accepting basic validity rules, there is no sense in which the philosophers themselves are thereby accepting any particular rules as setting the standards for legal validity.57 Secondly, in the scientific context there is an important self-consciousness about the limits of one’s knowledge and the reliance on others to do the important theoretical work that is clearly lacking in the legal situation. The highest court in the land does not believe itself to be relying on the theoretical work of philosophers in order to pick out successfully which laws are valid.

5. No Solution in the Background

The most natural response for a Searlean confronted by these problems is to say that officials’ dispositions to conform to the basic validity rule are a part of the “Background”58 capacities that support other institutional facts. Many of the points I have just noted about

57 See Hart’s distinction between internal and external claims about a legal system (1994, 102-03), his remarks more generally on the nature of legal theory in the Postscript (ibid.), and generally Dickson (2001). For criticism of this approach see generally Dworkin (2004).

58 Searle capitalizes the word to indicate that it has special meaning in his theory.
the basic validity rule do seem to reflect precisely the characteristics of the Background that Searle details: a disposition to behave in conformity with a constitutive rule, thereby collectively imposing the status function upon that rule, without the normally requisite mental representation of the rule. This would solve the problem of the officials’ needing to have the details of the basic validity rule in mind, or to be psychologically self-aware when using the basic validity rule in making their validity determinations. There might be some initial plausibility in this move since, as the most basic rule, the basic validity rule is an enabling condition for the determination of all other legal institutions within that system, precisely what we expect from Searle’s Background.

The Background is “the set of nonintentional or preintentional capacities that enable intentional states of function” (Searle 1995, 129). The Background explains how the constitutive rules (deontic powers) of an institution can cause conforming behavior in participants without them being directly aware of those rules (Searle 2010, 130). While the application of this idea is much broader in scope, Searle’s basic argument for it is linguistic, relying on the fact that we can only understand many words or phrases with multiple meanings and nuanced applications by relying on our possession of a Background network of “knowledge about how the world works [and]. . . abilities for coping with [it]” (1995, 131). What is true of language (itself a very basic institution) is then true for intentional states more generally. By enabling linguistic interpretation and conceptualization, the Background “structures consciousness,” includes and interacts with our “motivational dispositions,” and results in standing dispositions toward given kinds of behavior (Searle 1995, 132-37). These dispositions are “skills and abilities that are . . . functionally equivalent to the system of rules, without actually containing any representations or internalizations of those rules” (Searle 1995, 142). Instead, we are conditioned to behave in conformity with the rules that constitute our social and institutional structures by the responses of others. For Hart, what constitutes the evidence that a social norm is in place (and, for practice theory rules, partially constitutes the norms themselves) are the public responses of praise and blame (1994, 84). For Searle, these public responses condition the individual into conformity and collective acceptance (1995, 144-45).

The clear similarities between Searle’s description of the Background and the legal theorists’ explanations of the basic validity rule mean that this response is not to be swept aside lightly. However, there are serious problems in consigning the basic validity rule to Searle’s Background that would need to be answered. Recall the distinction between saying that a behavior is rule-described and rule-governed. Any behavior that is repeated over time and fits a particular description is a candidate for being rule-described. Habits are prime examples. Such rules do not themselves constitute reasons for the behavior as they are simply non-normative descriptions of the behavior and its regularity. The move to rule-governed behavior is precisely the point at which the rule itself provides a reason for the behavior. It is to see the rule as normative, calling for the behavior. 60

It is important for officials’ conformity with the basic validity rule to be rule-governed behavior. Legal officials must treat the fact that other officials conform their recognition

59 We will see that it is precisely this lack of “internalizations” of the rules that calls the Background into question as an appropriate description for conformity to the basic validity rule by legal officials.

60 The difficulty in explaining this move with respect to the rule of recognition as a practice theory rule is the basis of much criticism of Hart’s theory. See, e.g., Shapiro (2011, 102-04).
of valid law to the basic validity rule as a *reason* to do so themselves (albeit possibly implicitly or subconsciously). In Hart’s theory: “This consists in the standing disposition of individuals to take such patterns of conduct both as guides to their own future conduct and as standards of criticism which may legitimate demands and various forms of pressure for conformity” (1994, 255). It is all the more important for any theory that holds the basic validity rule to be duty-imposing (Shapiro 2011, 85; Raz 1979, 92).

Hart shares Searle’s starting point that legal rules can be valid (and hence institutional facts) when enacted and before anyone has occasion to conform to them. So when it comes to enacted law, legal theory seems to agree with Searle’s analysis that institutional facts take on a “life of their own.” However, for Hart, the basic validity rule is not such an enacted rule. Rather, “a legal system’s basic rule of recognition is treated as a rule constituted by the uniform practice of the courts in accepting it as a guide to their law-applying and law-enforcing operations” (Hart 1994, 258). The basic validity rule is therefore for Hart a practice theory rule, its normative force arising from its acceptance. Even if we reject Hart’s practice theory account of the basic validity rule (perhaps for leaving mysterious how we get from practice description to the normativity of a rule), it is at least clear that determinations of validity are rule-governed.

For the Background, however, Searle rejects the notion of rule-governance as too intentional. As mentioned above, the Background is the mechanism by which functions are ascribed to institutional facts non-intentionally. When behavior is rule-governed, the semantic content of the rule is causally determining the behavior (Searle 1995, 139). Searle’s example of speech acts makes it clear that this rule-governance can take place when an agent’s behavior is caused by an “unconscious internalization of the rules,” but this is still too intentional for the Background causal mechanism.

The difficulties in assimilating the basic validity rule to the Background should now be apparent. While the Background is an explanation of unconscious assignment of function, the basic validity rule must motivate a psychologically self-aware assignment of function for officials in its application. If the basic validity rule commands that validity determinations be made in accord with a foundational document like a constitution, the officials must be self-aware when they apply the constitution to make their validity determinations, even if they are not self-aware in applying a more basic validity rule directing them to look to the constitution. The officials might not be constantly self-aware of the criteria by which they determine legality (that is, the semantic content of the basic validity rule), but they are still making that determination self-consciously. That determination of legal validity is precisely the assignment of function Searle discusses. Certain

---

61 The description of the basic validity rule as a standing disposition might seem to buttress the claim that it is a part of Searle’s Background since Searle uses dispositional language to explain the Background. We will see, however, that the behavior that legal officials engage in (the determination of legal validity) as a result of these dispositions is too self-conscious and too foundational (in the assignment of function) for these dispositions to be part of Searle’s Background.

62 Hence legal rules generally are not the kind of rules that are described by his practice theory (Hart, 1994, 256).

63 “To the rule of recognition viewed in this way[,] the practice theory of rules is fully applicable.” Ibid.

64 See above, n.60.

65 “If we think of the Background intentionalistically, then we have abandoned the thesis of the Background. We arrived at that thesis in the first place only because we found that intentionalities goes only so far” (Searle, 1995, 140, denying that Background capacities like the rules of syntax can be understood as rule-governed). See also especially Searle (2010, 31).
statements “count as” law in certain contexts. Since the officials are making the assignment self-consciously and intentionally (even if the precise rules by which they do so are not fully and clearly formed consciously), the assignment of legal validity, and the internalized rules governing it, cannot be an element of the Background.

Searle explicitly denies an internalized rules explanation for the Background (1995, 145-47), and concludes that cases of Background causation cannot be described as rule-governed (1995, 140). For legal officials, it is clear that their behavior in determining legal validity is rule-governed, and that they have achieved any lack of self-awareness in applying those rules by internalizing them as they underwent their legal training. There is a right answer at which the officials are aiming in their validity practices, and they acknowledge mistakes as misapplications of their validity rules. Hence, while their acceptance (in Hart’s sense) still reflects a disposition to behave in conformity with the basic validity rules, and while this seems to be a kind of “background ability” that is suggestive of Searle’s analysis, the rule-governed, intentional, and conscious assignment of functionality in determining legal validity precludes locating the basic validity rule in Searle’s Background.

It might seem that this is just a problem of seeing the basic validity rule as a practice theory rule in Hart’s sense and that a solution would be to reject the practice theory and embrace Searle’s notion of the Background for basic validity rules. However, it would not be enough for a legal positivist merely to reject the practice theory of rules for the validity rule. In fact, Hart’s practice theory gets a lot closer to the Background than other candidates for the basic validity rule. In order to square with Searle’s Background, a legal positivist would have to give up on the idea that the basic validity rules are endorsed by or guiding the behavior of officials at all.

More recently, Searle has complicated the discussion by introducing notions of Background powers, which are norms that might come much closer to what we see with the basic validity rule. These do appear to be more prone to be explained as a form of rule-governance in that they set forth the content of those behavioral norms we might call customary or cultural for a given society, although they are generally still unconscious. They include, for example, forms of acceptable dress and speech, acceptable moral opinions, and other basic behavioral norms (Searle 2010, 155). Violations of these norms will typically be met with sanctions by other members of society, who will see violations as outside the realm of acceptable behavior and respond with expressions of disapproval or stronger censuring behavior (Searle 2010, 156-57).

There are two problems with this move for the basic validity rule. For one, Searle explicitly excludes legal sanctions from the kinds of pressure brought to bear on violations of the Background social norms (2010, 157). Of course, violations of the basic validity rule are not typically met with legal sanctions in the criminal or civil sense. If a violation of the basic validity rule is a mistaken validity determination by an official, then the usual response – a contrary declaration by officials of superior jurisdiction – may not count as a legal sanction. So, if we are narrow in our understanding of legal sanctions to exclude being overruled or perhaps losing one’s job, then this might not be a problem.

The bigger problem is that Searle is clearly talking about a background set of cultural norms that are enforceable by anyone and everyone (2010, 157). While ordinary citizens can make use of the basic validity rule themselves to make personal decisions about which laws they consider to be valid parts of the legal system they inhabit, their decisions do not have any legal weight. The basic validity rule is not aimed at them; it is
aimed at the officials who are charged with making determinations of validity (Shapiro 2011, 85).

Perhaps the most promising way out of this difficulty, offering the hope of seeing the basic validity rule as an element of Searle’s Background powers, is to see the officials as constituting a society unto themselves. This would be to say that the basic validity rule is a Background norm of a special club that officials belong to. The difficulty here is the hierarchical way in which officials are usually organized with respect to their validity determinations. That is, since Searle’s Background powers are possessed and enforced by everyone in society, there is a distributed and egalitarian character to their normativity that we do not see in the way the basic validity rule is addressed to officials. Even if we imagine officials at different levels of the legal hierarchy are equally important to the maintenance of the system by their use of the basic validity rule, that rule is still giving some officials the role of making fresh validity determinations and others merely the role of recognizing and applying the determinations made by others. One way out of this would be to see the “society” with the basic validity rule qua Background power as only consisting of those officials empowered to make ultimate determinations of validity in a given domain or jurisdiction. But as the basic validity rules are usually clearly aimed at and deployed by a wider set of officials, this move is artificial and runs counter to observed legal practice.

6. A Suggested Solution

Might we say that the basic validity rule is constitutive of (legal) institutionality in the same way we say that it is constitutive of validity? This would be to say that the basic validity rule is the constitutive rule by which the function ‘law’ is assigned to the declarations of legal officials. Applying this idea to Searle’s “X counts as Y in context C” formula, the ‘X’ term is the writings and pronouncements of legal officials; the ‘Y’ term is simply “law” or “legally valid”; the ‘C’ term is the context of the officials’ roles and duties (another institutional creation). The constitutive rule that assigns the ‘Y’ function to the ‘X’ term is the basic validity rule. That certainly seems to fit the analogous theoretical roles it plays in the two theories.

This shifts the focus of our discussion somewhat. No longer would we see a description of the officials’ validity practices as the ‘X’ term and something like “our (binding) criteria of validity” as the ‘Y’ term. That formulation generated the problem in the lack of a constitutive rule that could be codified by which the ‘Y’ function was assigned to the ‘X’ practice. In saying that the basic validity rule is itself only a constitutive rule but still an institutional fact, we are saying that there is something special and foundational about it, and perhaps similar foundational validity norms for institutionalized social practices.

This is an attractive solution but would require interposing an additional structure within Searle’s theory, providing a place for foundational constitutive rules that set the basic validity conditions for what might be called “institutional systems” like law. While Searle does provide for what might be called foundational institutional facts, he reserves that special role for language itself (1995, 60, 72-73), performing its symbolization function (1995, 60-61, 66). The details of this make it clear that it would not be directly available for the much less foundational notion of the basic validity rule of a legal system. Since symbolization itself is the function of language, we identify certain sounds
and groups of letters as instances of symbolization by linking the realization that thoughts represent and symbolize objects to the realization that language is the medium for doing so. To be such a foundational and self-identifying institutional fact is not to have the ascription of function set by a constitutive rule.

To entertain the possibility that the basic validity rule fits this description at a less foundational level, we would have to see it as self-identifying and not fixed by a constitutive rule, while it serves as the constitutive rule for generating the institution it validates. If we can interpose this additional intermediate foundation (not as basic as language), then the onus would be back on the legal theorists to fill in the blanks.66

There is something that remains mysterious about the way in which the mere description of other officials’ validity practices becomes a reason for action in conformity with those practices. Hart is content to rest upon the officials’ consideration of the practices as reason-giving, which they internalize in seeing it as a basis for criticism, doing so for a wide variety of possibly personal reasons (1994, 55-57, 83-91, 108, 203). But if we can say that the basic validity rule is self-identifying in that the very description of officials’ validity practices somehow carries with it the demand for conformity, then we would not expect it to be fixed by constitutive rule. It would itself be the foundational constitutive rule for the institution of law and a search for its constitutive rule would be as misplaced as the question of whether it is legally valid.67

While this does not entirely solve the problem of explaining legal normativity, it at least helps us to narrow down the issue by showing that legal normativity arises out of the demand for conformity that is carried by the basic validity rule as a self-identifying constitutive rule. Whatever makes the basic validity rule binding upon officials is then also the source of legal normativity more widely. It certainly seems reasonable to say that officials are under a legal duty to adhere to the basic validity rule in their determinations of validity. Since that rule then determines which other legal norms are legally valid, it gives the officials the deontic power to impose novel legal norms in conformity with its dictates, transmitting normativity to the legal requirements they create and apply.

7. Conclusion

The basic validity rule that sets forth the criteria by which the elements of a legal system are to be recognized as valid cannot fit neatly into Searle’s picture of institutional reality. It must be an institutional fact rather than a simple social fact because the rule counts as something more than the bare fact of the validity practices of officials. It now carries a demand for compliance by legal officials. The description has become a reason to behave in certain ways.

It cannot be merely an informal institutional fact because it is a standard by which other validity decisions can be deemed incorrect (a standard by which counterfeit instances can be recognized). Furthermore, it is not flexible and spontaneous, characteristics of informal institutions. Yet it cannot be a formal institutional fact because it cannot be codified in Searle’s sense. There is rarely any self-aware official recognition of its

66 One possible answer to this challenge (although not made in response to such an alteration of Searle’s theory) is suggested by Marmor’s notions of “deep conventions” (2006a, 363-71; 2007, passim).

67 For an analysis of the relation of the validity practices to the officials’ reasons for accepting them as binding, see Dickson (2007, 4-9).
status. We do not see official fixation of the characteristics of its type by which we could decide which among token candidate basic validity rules are valid. (It is constitutive of legal validity itself.) There is no constitutive rule by which the validity practices of officials are deemed to function as normatively binding. Finally, there is no official, explicit act institutionalizing the basic validity rule.

This leaves us in a conundrum when trying to understand law from within Searle’s institutional framework. One possible solution is to interpose a new class of foundational facts for institutional systems. One result of this would be to refocus the jurisprudential quest on explaining how descriptions of officials’ validity practices carry a demand for conformity, which is really at the root of the questions about what (if anything) legitimates legal authority.68

References


68 The author would like to thank Detlef von Daniels, Julie Dickson, David Enoch, John Gardner, Michael Giudice, Les Green, Dennis Patterson, George Pavlakos, Cindy Phillips, Veronica Rodriguez-Blanco, John Searle, James Specyal, Amie Thomasson, Jeremy Waldron, and Ekow Yankah, as well as the anonymous reviewers for this journal for their helpful comments and discussions.


