

Why a Theory of Institutions cannot include both Rules and Norms

John Joseph Wallis

February 2019

This paper has appeared under several titles over the past three years. Most often as “What Institutions Are” with a variety of subtitles. I owe thanks to many people, prominently to Price Fishback and Sumner Lacroix, to Lee and Eric Alston and Tomas Nonnemacher, to Naomi Lamoureaux, to Christopher Morris, Peter Murrell, Ethan Kaplan, and Roger Betancourt, and to my graduate economic history reading group and students in my graduate seminar. A complete list of seminars and acknowledgments will be forthcoming soon.

Abstract: How institutions should be defined has always been an unsettled area of institutional economics. North defined institutions as the rules of the game, including both formal rules and informal rules (norms, conventions, and self-imposed constraints) as well as their enforcement characteristics. A debate within the institutional economics community over defining institutions-as-rules versus institutions-as-equilibria has raised serious questions about what institutions are and what we are trying to explain. Unfortunately, rules have been poorly defined. In North's definition laws are rules and norms of behavior are both rules. One is formal and the other informal. How can they both be rules? This paper argues that institutions should be defined as "agreed upon rules," where agreement is a deliberate collective action by two or more people. It explores dimensions of agreed upon rules normally ignored in the institutional literature, most importantly that many rules are default rules that are not prescriptively enforced. Norms of behavior do not require agreement, even though they do exert a powerful influence on individual choice. The substantive difference between rules and norms that matters conceptually for how we understand institutions is that rules require agreement, norms of behavior do not. Oddly, norms are always followed or they are cannot become norms of behavior, while rules are often not followed. The distinction between formal and informal is natural, perhaps, but leads to confusion rather than insight. Any consistent "theory of institutions" cannot encompass agreed upon rules and norms of behavior because they are fundamentally different social processes.

“Institutions are the rules of the game of a society or more formally are the humanly devised constraints that structure human interaction. They are composed of formal rules (statute law, common law, regulations), informal constraints (conventions, norms of behavior, and self imposed rules of behavior); and the enforcement characteristics of both.” (North 1992, p. 4).

*

You may not have noticed it, but a debate has been brewing about what institutions are. It is a tempest in a small teapot for most economists, perhaps even a debate about the number of angels dancing on a pinhead, but it won't remain so. Development agencies spend billions of dollars a year thinking about institutions and trying to improve them in their client countries. The ability to implement findings discovered by randomized controlled trials depends on the quality of institutions in the recipient country. Imagine the confusion if economists do not know whether institutions are agreed upon rules, which might be implemented by a government; norms of behavior, which no one knows how to change, except perhaps through advertising and propaganda; or beliefs about the world that individuals hold? Current definitions of institutions often include all three, but not all do. Not only do we not know the difference between rules, norms, and beliefs, we are on the verge of reading agreed upon rules out of the institutional canon because rules are often not followed. We have been accumulating evidence, reviewed by Alessina and Giuliano (2015), that norms and/or culture often exert a strong influence on behavior, even at a distance in time and space. Alessina and Giuliano suggest that we separate institutions and culture, defining institutions exclusively as formal rules. More generally, if we adopt the idea that rules only matter if they are followed, and we can show which rules are or are not followed econometrically using well identified empirical specifications, then we can leave all the “unexplained” patterns of behavior to norms, culture, and beliefs. If we do that, we are

making a serious mistake. We do not understand rules as well as we should and, as a result, we do see how they work in human societies. Agreed upon rules are often not prescriptively enforced, they are “default” rules. The institutions-as-rules versus institutions-as-equilibria debate only deepens our misunderstanding of agreed upon rules.

Although North usually added a qualifying “enforcement characteristics” clause to his definitions of institutions as the rules of the game, much of the disagreement about what institutions are revolves around enforcement. In North’s two part division of institutions into formal and informal rules, one may conclude that the characteristic enforcement of formal rules was external third party enforcement, often by a government. Informal rules could not be enforced by an external third party, however, indeed how they were enforced was a major puzzle. Advances in game theory led to insights about how expectations of the behavior of others acted as constraints on individual choices. If you have beliefs about how others will behave or how they will react to your behavior, those beliefs conditions your choices. Norms of behavior are patterns of behavior. If enough people see a pattern in the external world and their response conforms to the pattern, then a “self-enforcing” norm of behavior can emerge. A norm of behavior can be a self-enforcing equilibrium, and institutional analysis using these kinds of ideas has come to be know as institutions-as-equilibria (Greif and Kingston, 2011; Hindriks and Guala, 2014).

The equilibrium approach solves the problem of enforcing informal rules like conventions and norms of behavior, by locating the source of enforcement in the internal incentives determining individual choice. But that leaves the enforcement of formal rules hanging. People had long thought about how rules could be enforced through coercion, the

threat of violence. Hobbes built his theory of government and law on the willingness of individuals to concede control of their capacity for violence to the direction of the Leviathan, in return for which Leviathan was incentivized to enforce his rules and provide social order, otherwise people would not follow his rules. Hobbes makes one of the first institutions-as-equilibrium arguments in the social sciences. Hobbes's rules and laws are prescriptive: the rules describe behavior and punishments for deviations from the behavior.

The central argument of this paper builds on the insight that many agreed upon rules are not prescriptive rules. Agreed upon rules result from a deliberate act of agreement between two or more people. Agreed upon rules may apply to many others who were not part of the agreement. The agreement in agreed upon rules does not imply consent from those to whom the rule applies. The form of the rule describes a behavior that individuals *may* follow, but if they do not follow it the rule is not actively enforced. These are *default rules*. Default rules are not prescriptively enforced. They only apply if two or more parties to a relationship ask the third party, perhaps a court, to enforce the rule. It is not that the cost of enforcing default rules is so high that they are only applied in high value cases, it is simply that no one, no third, second, or first party is out in the world enforcing those default rules. Default rules typically apply to relationships between people. When two individuals are in an appropriate type of relationship, a default rule may apply to them both. If one or the other decides to invoke (or access) the default rule, then the appropriate third party makes a decision about the rule, and if necessary enforces the decision by the agreed upon means of doing so.

For example, in many states in the United States married couples, one or both of whom seek a divorce may, if they can agree, write their own divorce agreement within wide boundaries

(arrangements for children being the main exception). If the divorcing spouses cannot agree, the courts impose a pre-arranged set of rules determining property division, alimony, and the like. The rules the court will enforce are default rules. The court will not impose the rules unless one, or both, of the parties asks the court to apply the rules. In the words of Mnookin and Kornhauser (1979) divorce takes place in the “shadow of the law.” Default rules cast a different kind of shadow from prescriptive rules. This aspect of rules has not been appreciated by the institutional literature.

North’s definition of institutions was broad and inclusive. He wanted to understand how non-price constraints affected individual choice. Every aspect of the external social world, and a few of the internal world (habits), entered into his definition. His approach was rooted in a methodological individualism, where if the choices of individuals could be understood, the problem of understanding how society worked would largely be solved. North wanted to build a general theory of institutions and institutional change, with the key insight built around how institutions constrained individual choice. This is the conceptual intuition behind “institutions-as-rules” as North developed it. In his framework, *norms and conventions are rules*.

The contribution of the equilibrium approach was to make crystal clear that how individuals interacted was as an important determinant of the decision individuals made. Individual decisions were inseparable from the expectations of how people around them would behave. Individual decisions could not simply be aggregated up to social outcomes. As the sociologist Emile Durkheim noted, the choices and behavior of individuals creates “social facts,” patterns and realities in the social world around us. Phenomena that no individual intended to

create but which, nonetheless affect everyone.¹

The truly brilliant insights of the equilibrium approach paint a picture of the world where most social behavior is governed by patterns of self-enforcing norms of behavior. If norms are rules, and equilibrium analysis shows us how those rules are enforced, where is the corollary enforcement of agreed upon rules? Norms of behavior are always followed by enough people to make them norms of behavior. If too many people don't follow the norm, *it cannot be a norm because it will not be self-enforcing*. If rules that result from deliberate agreement are not followed, implying that they are not or cannot be enforced, does that mean the agreed upon rule is not an institution?

Now we are at the crux of this paper. If norms are rules, and agreed upon rules are rules, why are all norms enforced but many agreed upon rules are not? Why are some rules followed and others not? Don't we need to understand why people follow rules to answer these questions? That is the position of the institutions-as-equilibrium approach. These are good questions, and we can't answer them if we assume that all rules are either enforced or they are ineffective. If agreed upon rules that are rarely enforced, rules that specify behavior and then behavior does not follow the specification, can they affect how societies work? If the answer is yes, then there really is a difference between norms of behavior and agreed upon rules, but the

¹“The core idea in the institutions-as-equilibria approach is that it is ultimately the behavior and the expected behavior of others rather than prescriptive rules of behavior that induce people to behave (or not to behave) in a particular way. The aggregated expected behavior of all the individuals in society, which is beyond any one individual's control, constitutes and creates a structure that influences each individual's behavior. A social situation is ‘institutionalized’ when this structure motivates each individual to follow a regularity of behavior in that social situation and to act in a manner contributing to the perpetuation of that structure.” Greif and Kingston, 2011, p. 25.

difference is not one that the institutional literature assumes. Norms are always followed by enough people to make them norms of behavior. Many agreed upon rules are often not followed, and there are good reasons why they are not that do not involve the costs of enforcement. Default rules are more powerful coordinating tools if they are not actively enforced. The institutions-as-rules approach lacks a theory of enforcement comparable to the institutions-as-equilibria approach. Nonetheless, the reality of how rules work in the world us requires that we have two different theories about different institutions: one about agreed upon rules and the other about social norms, culture, and beliefs. We cannot have an integrated theory of institutions that attempts to explain agreed upon rules and norms with the same conceptual model, as if rules and norms are just different aspects of the same fundamental social process. They are not. This raises the question of whether we want to call both agreed upon rules and norms of behavior institutions? If we want to call them both institutions, then we have to abandon the idea that we will have a unified theory of institutions.

This does not mean either norms or rules are not important, they are both critical elements in understanding how societies work. But they are not they same thing and we will not understand how they interact if we insist on the same conceptual framework for understanding both.

The next three sections of the paper examine default rules in greater detail. They show how default rules enhance social coordination and the value of relationships even though they are not prescriptively enforced. A way of picturing the universe of rules and organizations in a society, the rule environment and organizational ecology, is introduced as a way to keep track of how rules and organizations interact with one another. The third section considers the role of

primary and secondary rules in the rule environment. The fourth section provides detailed examples from the literature. The fifth sections return to the critique of institutions-as-rules made by the institutions-as-equilibria approach and reevaluates the substantive differences when we better understand how agreed upon rules are, or are not, enforced. The final section extends the insights to governments, collective choice, and the nature of institutional change.

**

Some definitions will help:

Rules: Institutions are “agreed upon rules.” This is a narrow definition of institutions and of rules. Rules result from a deliberate agreement between two or more individuals. While we may all agree that the sun rises in the east, that kind of “agreement” does not underlay an agreed upon rule. Agreements are deliberate, the result of a collective choice. The scale of rules varies from families on the small end to nations and international organizations on the large end. When we speak of the “institutional structure” of an economy we speak about the structure of agreed upon rules. This parallels the distinction between institutions and culture in Alesina and Giuliano, where institutions are “formal” rules and culture is enduring patterns of values and beliefs.² The definition of institutions as agreed upon rules obviates the distinction between

²“On the empirical side, most papers (if not all) follow the definition adopted by Guiso, Sapienza, and Zingales (2006), where culture is defined as “those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation.” Empirical papers, therefore, combine values and beliefs in the same definition.

On the theoretical side, values and beliefs are often treated differently. Several authors have developed models in which *culture* means beliefs about the consequences of one's actions, but where these beliefs can be manipulated by earlier generations or by experimentation.” Alesina and Giuliano, 2015, p. 900.

“formal” and “informal” rules, since agreed upon rules within a family, that apply only within that family, surely would be regarded as “informal,” but those rules are institutions just like a national constitution. What matters is that all rules are agreed upon rules, no matter what their scale of origin or application.

Organizations: Organizations are groups of people that agree upon rules that structure their internal relationships. Organizations may share a common purpose across their members, but that cannot be a requirement for an organization since many large organizations include members with competing purposes. How organizations agree upon rules varies across organizations. All organizations have “primary” rules that govern relationships as well as “secondary” rules that determine how new rules are made and existing rules are altered.³ The existence of an agreed upon rule *does not* imply consent by the all members of the organization (either active or passive) to the rule. The secondary rules allocate the ability to make decisions about to certain individuals, or to individuals occupying certain positions in the organization, and perhaps following certain procedures. Individuals can decide to be in or out of an organization, but that choice does not create “consent” in the active way that we mean by “agreement.” Individuals recognize that if they belong to this organization, certain rules apply, and that they may have little or no influence over the rules.

How do rules and organizations fit together? Seeing how default rules that are not always enforced nonetheless enhance coordination in societies and value in relationships is easier if we think about the interaction between multiple rules and multiple organizations. The collections of rules and organizations will be called rule environments and organizational

³The concept of primary and secondary rules originates with Hart, 1961/2012.

ecologies. The terms represent systems of rules and populations of organizations and their interactions that we can see visually, verbally, and conceptually. When multiple rules and organizations interact, some rules will be default rules: rules that exist but are not always applied or enforced. In a rule environment with many organizations it is unlikely that all the rules in all the organizations will agree with one another. If two rules are “incongruent,” if the behavior specified in one rule conflicts with the behavior specified in another rule, then both rules cannot be precriptively enforced. But incongruent default rules can co-exist in the same rule environment.

When I was younger, I spent several years working construction jobs. I was a member of the Laborer and Hod Carriers union, which was the lowest rung of the construction unions. There was also a Carpenters Union. The construction firms I worked for used union labor, and I often worked as a carpenter’s helper. The relationship between any carpenter and his helper (me) is always unique. Each carpenter has his or her way of doing things, and each laborer is more or less skilled at different tasks. Each laborer/carpenter pair reaches an agreement about how they will work together. This agreement results in their “agreed upon rules” that apply at the level of the laborer/carpenter pair. We can accurately, if trivially, say that each laborer/carpenter pair is an organization. Their agreed upon rules are an institution, but an institution that only governs their relationship, what we can call “internal” rules, rules that apply only within an organization.

The laborer/carpenter pairs are embedded in three other levels of organization: the firm they both work for, the two unions whose work rules affect their relationship, and the state government whose health and safety rules affect their relationship. In each case, the rules at the

other “external” levels of organization affect the productivity (positively or negatively) of the relationship between a carpenter and his helper, and so affect the productivity of the construction process. Each laborer/carpenter pair is embedded in a larger “organization ecology” as shown in table 1. While every organization has its own unique internal rules, all the rules in the system potentially interact.

Organizations:
The State Safety Commission
The Unions:
Carpenters Union and Laborers and Hod Carriers Union
The Construction Firm
The laborer/carpenter pair

Employment on a union construction site involves lots of rules. On a big project union representatives and state safety officials make regular visits to check that the rules are being followed. And yet, many of the rules are not followed to the letter. Viewing rules, or laws, as strictures that *must be followed* gets in the way of understanding how rules can increase the value of relationships by changing the outside options available to the parties. One organization’s rules can serve as external rules for other organizations. The interaction of the rules across organizations contains a very important aspect of rules that we often do not appreciate: not all the rules are followed, and *not all the rules can be followed*, because the rules at different levels

are not always in agreement with one another. When rules and the behavior under them (or other rules in the environment) do not match we will say that the rules are “incongruent.” In the laborer/carpenter pair, the rule the pair reach about may be incongruent with the rule that the unions enforce.

The simplest example is hammering (driving) nails. It is very clear from the union rules that carpenters are supposed to drive nails, laborers are not.⁴ Laborers always carry hammers, however, and laborers are often asked to drive nails. Why is there a clear union rule prohibiting laborers from driving nails when it is equally clear that they often do? When a laborer drives nails it is not thought of as *breaking* the rule, even though everyone involved understands the

⁴The difference between hammering a nail and driving a nail is only obvious to people who know how to drive nails. On a construction site you are expected to “drive” home a nail with one or two blows of the hammer, not the dozen or so that an amateur requires. Even laborers who do not drive nails carry claw hammers, because an important function of claw hammers is to remove nails. Laborers spend a lot of time taking nails out of wood. My first construction job was not a union job and I had to quickly learn how to drive a nail. Obviously, later on when I joined the laborer’s union and started working with a union carpenters, whether I would be allowed to drive nails or not depended in part on my ability to do so.

Table 2
The Rule Environment of
Laborers and Carpenters
with respect to Driving Nails

Organizations:	Rules:
The State Safety Commission:	Rule I: Don't drive nails without wearing safety goggles
The Unions:	Rule II: Carpenters drive nails, laborers do not drive nails
The Construction Firm:	Rule III: Do what the carpenter says
The laborer/carpenter pairs	Rule IV: Whatever arrangement the carpenter and laborer work out between themselves determines who drives nails

union rules about driving nails. The “Rule Environment” for driving nails is shown in Table 2.

Every laborer/carpenter pair works in this rule environment. The union rules give the carpenter the formal authority to drive nails. The construction firm agrees to the union rules, and all of the carpenters belong to the carpenters union and all of the laborer's belong to the laborers union. The carpenter always has the option of driving the nail. Laborers do not always have that option because they do not possess the authority to drive nails whenever they want to, while carpenters can. We might say that carpenters have the power to drive nails, and laborers do not. What this really reflects is that there is an external rule allocating authority over the decision about who drives nails to the carpenter. The nail rule a laborer/carpenter pair agree to can be incongruent with the union nail rule. The union nail rule is a *default rule*. It is always there, but only applies if either the laborer or the carpenter disagree and decide to access the rule. They can reach an agreement negating the union rule in their relationship, but they cannot make the

union nail rule go away. That is critical to how default rules enhance coordination in relationships. Invoking the nail rule is always an outside option for the carpenter or the laborer. Whether a laborer drives nail or not depends on the relationship between individual carpenters and laborers. Some carpenters never let laborers drive nails, other carpenters do. Even within the same firm different pairs of carpenters and laborers have different agreements, the agreed upon rules, for driving nails. Laborers and carpenters can enter into relationships with each other in which either drives nails, but carpenters cannot credibly promise not to drive nails, since carpenters always possess the authority to drive nails. If there is a dispute between laborers and carpenters that has to be resolved by management or the union, according to the formal rules, carpenters drive nails.

The nail rule seems pretty trivial until you think about it for a moment. At any specific firm, laborers and carpenters are transitory. Since new laborer/carpenter pairs are created quite often, by giving the carpenter discretionary authority from the beginning, new partners don't have long involved discussion about how they should work together. The laborer starts out doing what the carpenter says. But relationships grow and change. No construction manager, however, wants to resolve a dispute between a disaffected laborer who goes to the boss with a complaint that his carpenter agreed to let him drive nails yesterday, and now the carpenter has reneged on their agreement. Disputes of that sort just do not arise. Relationships grow and change within the framework of the external rules. All kinds of coordination within the construction firm is ordered by the rule allocating the authority over driving nails to the carpenter, as well as hundreds of similar rules. The rules also increase the productivity of every laborer/carpenter pair, no matter what agreed upon rules they actually reach in practice, because

the default rules allow individual laborer/carpenter pairs to work out the arrangement that works best for them.

The productivity and profitability of a construction firm depends on the productivity and effectiveness of its work units: the laborer/carpenter pairs. The rules that the unions and the safety commission create are important part of the external rules that the firm and the laborer/carpenter pairs can access *if they want to access them*. The driving nails rule allocates authority over nails to the carpenter, but *these default rules are not followed in practice*. It works bests to give carpenters the authority to drive nails, but it also works best at times to let laborers drive them too.⁵ Heterogeneity can be good. Letting each laborer/carpenter pair work out their best arrangement for working together makes the firm more productive.

It is very important to understand that the union nail rule *only works because it is not always followed*. The rule provides a framework for relationships between the firm and its employees, and between the laborer/carpenter pairs. By allowing both the firm and the laborer/carpenter pairs the option of invoking the nail rule, the rule makes the laborer/carpenter pairs and thus the firm, more productive and profitable. The opportunity to use rules in this fashion only exists if the rules are incongruent, if the rules across the rule environment do not agree with one another completely. If there is a dispute between the laborer/carpenter pair that a firm's managers must resolve, or a dispute within the firm that makes it to the union level, the

⁵Aghion and Tirole show how allowing a difference between formal (union) and real (laborer/carpenter pair) rules that allocate authority can increase the value of the firm. Indeed, they show that sometimes the best formal rules are not only rules that are not followed, but sometimes the best formal rules are rules that differ from the actual rules people follow. I will come back to Aghion and Tirole in the next section, but not adopt their language of formal and real rules.

dispute is resolved by acknowledging the carpenter's authority to drive nails. Since everyone knows how the rule will be enforced, those disputes may occur, but disputes rarely rise to the level of a "third-party" decision and enforcement of the rule. The idea is that the unions will always rule in favor of the carpenter. That is, that while the union rules distinguish between the categories of laborers and carpenters, they will always apply the nail rule impersonally, the same to every laborer/carpenter pair. Having a union rule that is incongruent with the agreements laborer/carpenter pairs reach actually reduces disputes between laborer/carpenter pairs and enhances coordination and productivity.

Suppose, instead, that enforcement of the union nail rule depends on the identity of the parties involved. Unions often value long tenured members. Suppose a highly experienced laborer is able to bend the enforcement of the nail rule in his relationship with an inexperienced carpenter. The unions might not be willing to enforce the rules against a proven experienced member. If so, the carpenter's authority becomes weaker, in the sense that the third parties who enforce the rules are unwilling to side with a specific carpenter against a specific laborer. This is bad news, both for the firms and for the laborer/carpenter pairs. The firm now needs to know more about the identity of each laborer and carpenter, and each laborer/carpenter pair needs to spend time figuring out which outside rules apply to them. The productivity of the laborer/carpenter pairs and of the firm will be lower if the rules about driving nails at the union level are identity rules, that depend on the personal characteristics of the laborer/carpenter pair, instead of impersonal rules, that treat all carpenters the same and all laborers the same. This is true about any relationship that draws on an external rule to help structure an internal agreement within an organization.

The gist of the example is that a relationship between a laborer/carpenter pair can be made more valuable if the internal rules the pair adopts are embedded in the context of external default rules that the pair can access, and the external rules will create more value if they are impersonal rules than if they depend on the identity of the people involved. The logic of the employee/employee relationship in the nail example extends to employee/employer relationships in general, to union/employer relationships, and to state/private organizations relationships all of which can use default rules in this way. In the wider world, relationships between borrowers and lenders, between managers and shareholders, between contracting parties of all types, between teachers and students, and an innumerable number of relationships are all affected by the existence of external default rules and their enforcement.

I am not arguing that impersonal rules are always better rules than identity rules. I am arguing that in the circumstances that apply in the nail rule example, which apply to a wide variety of relationships in human societies, that the logic of the nail rule applies. There are many circumstances where the an actively enforced prescriptive rule, like traffic rules, is more efficient than a default rule. There are many cases where an identity rule is more efficient than an impersonal rule, if only because the identity rule can be enforced where the impersonal rule may not be.

I am arguing is that the transaction costs of coming to agreements and enforcing them within relationships can be significantly reduced by the existence of external default rules that the parties can access at their discretion. Moreover, I am arguing categorically, that the transaction costs of using impersonal default rules to order relationships will always be lower than the transaction costs in using identity default rules, because identity rules always require

parties in a relationship to know the identity of the other party before they know how the default rule will be applied.

The nail rule gives us a clear example of how a rule that is not enforced and whose prescription may or may not be followed by people to whom the rule directly applies, can nonetheless exert a powerful influence over how people behave in relationships. The implication that impersonal rules are more effective coordinating institutions than identity rules has important consequences for how we understand modern development, both political and economic, as impersonal rules do not become widespread and credible until the middle of the 19th century, sometime in and after the 1840s (Wallis, 2005 and 2006; NWW 2009; Lamoreaux and Wallis, 2017 and 2018).

The notions of a rule environment and organizational ecology are important concepts in their own right and will be particularly useful when we return to consider the difference between norms and rules in the fourth section. The concepts have a depth that the nail rule does not illustrate, but is important to understand. I have made two distinctions between rules so far, between prescriptive and default rules, and between identity and impersonal rules. The third and last distinction also matters: the difference between primary and secondary rules.

In his classic book *The Concept of Law*, H.L.A. Hart analyzes the requirements for a system of law to exist. His target are conceptions of law as commands, the embodied will of a sovereign power.⁶ Hart gives the sovereign a name, Rex, and asks whether the mere will of Rex

⁶Historically the analytical approach to law as command was developed by John Austin. These ideas are still prominent in the 21st century, as in the work of Ronald Dworkin.

is sufficient to create a system of law. His answer is no. A naive and simple interpretation of Hart's logic and intuition begins with Rex as the lawgiver, but in the context of courts and organizations that implement the laws Rex commands. What happens if Rex wakes up one morning and decides to change the laws? How do those who administer the laws know what the new laws are? How do they recognize the new laws? Hart calls the rules for making new rules or changing old rules "secondary rules." They are rules of recognition, a prerequisite for a system of rules with more than one organization involved.

Secondary rules are more than simple rules of recognition, however. Secondary rules fix how rules can be created and changed, but they also create conditions that rules must reach. In the 1830s and 1840s when the states in the United States began changing their constitutions to require that legislatures pass "general" laws for specific functions, like chartering corporations. General laws applied equally to everyone, their opposite were special or local laws that applied only to an individual or small group. At that point the secondary rule became that no law that was not a general law could be a law. This is more than a rule of recognition, but it is clearly a secondary rule in Hart's framework. The rule environment and organizational ecology depicted in Table 2 is an incomplete description of the complete rule environment which would include secondary as well as primary rules. The organization listed in the first column and the primary rules listed in the second column are embedded in a set of secondary rules that determine how rules can be made, by which organizations they can be made, and the process by which they can be made. The secondary rules are an integral part of the rule environment and organizational ecology of a society.

Can secondary rules be norms or must they be agreed upon rules? This is a tricky

question to answer, but the answer will be important when we come to determine whether both norms and agreed upon rules can be treated under the same theoretical framework. The United States Supreme Court has the authority to declare laws passed by Congress and signed by the President as “unconstitutional.” Most Americans believe that the Constitution gives the Supreme Court the right to declare laws unconstitutional, but you will search in vain for such a grant of authority in the text of the Constitution or its amendments. Does this mean that the Supreme Court’s power is a norm? Or that it is an informal rule? No, it doesn’t. The power of the Supreme Court is the result of explicit and deliberate agreements about the relative power of the Court, the Congress, and the President that have been contested but reached over the last two centuries. It is impossible to read the history of the law in the United States any other way, but that does not mean the power to declare laws unconstitutional is written in the Constitution of 1787.

School children in the United States learn “how a bill becomes a law:” the secondary rules requiring that identical legislation be passed by the House and Senate and then signed by the President before it is recognized as a law. Those secondary rules that play a central role in both the legal system of the United States and the institutional structure of agreed upon rules. The institutional structure, however, is far more complicated and rich than the legal structure. All organizations in the United States have agreed upon rules that structure the internal relationships between members of the organization. There are also rules in the rule environment that govern relationships between organizations.

The richness and connections runs deeper. Within the rule environment and organizational ecology what is a primary rule at one level of organizations may be a secondary

rule at other levels of organizations. The chartering of corporations provides a particularly apt example, because corporate charters are the rules for forming an organization. During colonial times, the British government had the power, a secondary rule, in the form of the King and Parliament to create corporations. Each corporation was unique and special. Corporations included the Commonwealth of Virginia as well as the Bank of England. At independence the states assumed the power to create corporations, and they began to do so in large numbers. In 1783, the colony of New York passed the first general incorporation law for churches, than allowed any group of citizens to form a church with a corporate identity. By the 1850s many states amended their constitutions to include provision requiring state legislatures to create general incorporation laws and, in some states, banned special charters.

At the level of the state government, the constitutional requirement that the legislature pass general incorporation acts was a secondary rule binding the legislature. A unique corporate charter issued by the legislature could be challenged in the courts, and they were, for creating a corporation in violation of the secondary rules that required general laws that apply to all corporations equally. The general law provisions in the state constitution were secondary rules for the state legislatures. The general incorporation laws the legislatures passed were primary rules for the state legislatures, those laws governed behavior directly. For the corporations formed under the general laws, however, the state's general law was a secondary rule. Corporations were free to form, but they could not violate the terms of the general law when they wrote their charters without laying themselves open to charges that their charter was in violation of the law. State general laws often regulated how shares in a corporation would vote on corporate decisions. The actual corporate charter written was a primary rule for the corporation

(that was true whether the charter was based on a general law or was a special law itself).

Within the organization of the corporation, however, the charter was a secondary rule that specified who in the organization could make decisions, about what things, and following what procedures. The rule environment contained a pattern of rules, descending from the constitution to the legislature to the individual corporation in a kind of rule ladder that looks like:⁷

(1) Secondary > Primary > Secondary > Primary > Secondary > Primary

If we were to write out the rule environment and organizational ecology for corporations in a typical American state the primary and secondary rules would connect in this way. The same rule may be a primary rule at one level of organizations and secondary rule at other levels. Moreover, in the case of American state governments and the corporations they charter, almost all of the rules are default rules, not prescriptive rules. States do not check to make sure corporations are following their charters. The charters are agreed upon default rules. Corporations and their managers, stockholders, creditors, debtors, and employees use the charters and the general acts as outside options in their relationships. The value of those relationships is enhanced by impersonal default rules, if the same rules apply to all corporations and all the managers, stockholders, creditors, debtors, and employees. That is true even if nobody follows the default rules. I am sure you can now see that “not following” the default rule does not mean that no one’s relationships conform to the rule, it means that many relationships do not and no one is “constrained by the rule” to adopt a relationship that follows the agreed upon rule.

⁷I use the word ladder rather than a hierarchy because, as I will show in the next section, it is not clear that rules at lower levels of organizations cannot be default rules that enhance coordination at higher levels of organizations.

The last point to emphasize about the rule environment and organizational ecology and the way rules can be simultaneously primary and secondary at different levels of organizations is that these must be agreed upon rules, they cannot be norms of behavior. Almost no one living in the United States, aside from some corporate lawyers, a handful of economic historians, a perhaps a few corporate finance specialists understand that anything like the primary/secondary rule ladder exists. It does not exist because of behavioral patterns, it exists because of deliberate agreements. You could not infer what the primary/secondary rule ladder looks like by studying just the behavior of corporations, because even corporations chartered under the same restrictive general law do not operate in identical fashion with respect to behavior that is clearly laid out in the general incorporation act. Managers and shareholders that reach an agreement within a corporation that violates the charter, will persist unless some manager or shareholder decides to take the arrangement to court. In most cases those kind of disagreements will be settled within the corporation, just as disputes between laborer/carpenter pairs are not taken to the union. As a result, you cannot infer the laws from observing behavior, because behavior need not follow the law. These are default rules.

Just as with the unique relationships between laborers and carpenters, if every corporation adopts different patterns of behavior, the norms cannot be built upon behavior that does not exhibit a pattern and generate expectations of behavior. Even if the laborer, carpenters, and corporations adopt “informal” rules to govern their internal relationships, those informal rules cannot be norms if they vary from pair to pair or corporation to corporation. What matters is not whether rules are formal or informal, what matters is whether rules are agreed to. Norms are not agreed to. We do not want to put all “informal” rules into the category of norms. Instead,

we should drop the formal/informal distinction.

This section presents some concrete examples of default rules. One of the concrete examples is abstract and theoretical taken from Aghion and Tirole's paper "Formal and Real Authority in Organizations." Another is taken from Lisa Bernstein's study of relational contracting in the cotton industry. The third is from the empirical evaluation of no-fault divorce rules by Mnookin and Kornhauser and others. The examples all show how default rules work to enhance the value of relationships. In all three cases, default rules work better as coordinating tools if the rules do not correspond to what the parties to the relationships normally do, that is, if there is incongruity between the default rules and the agreed upon rules within the relationships.

Aghion and Tirole are concerned with how rules (and authority) affect the value of a firm. They contrast "formal" authority and "real" authority within organizations. For this example I will accept their terminology, although for reasons I have already talked about, the distinction between formal and informal is problematic. Aghion and Tirole want to explain why an organization, like a business firm, might structure a formal rule that applied within the firm when the firm did not, in practice, enforce the rule. In other words, Aghion and Tirole asked why a firm might create default rules that operate only within the firm, with no appeal to an external third-party enforcer, and then not prescriptively enforce the rule within the firm?⁸

⁸"This paper develops a theory of the allocation of formal authority (the right to decide) and real authority (the effective control over decisions) within organizations, and it illustrates how a formally integrated structure can accommodate various degrees of "real" integration." Aghion and Tirole, 1997, p. 1. My discussion of their paper is much simpler, and probably simple minded, than their analytical exposition, but I think it captures the main insights of their paper and translates it into the larger issues that this book is concerned about.

Aghion and Tirole model the relationship between a principle and an agent who must make a decision about which of many projects to pursue.⁹ I am going to change their example a bit to make it more tangible, using the example of a firm where upper management sets the rules for the working units. A lower level manager in charge of a work unit is responsible under the upper management level rules for making a specific decision about where to locate a new store. I'll call those rules the "firm's rule." The lower level manager heads a work unit of employees who gather information about potential locations. The rule that the "manager makes the decisions" is external to the work unit. It is like the nail rule that carpenters drive nails, laborers do not. In practice, the manager usually consults with his subordinates and often allows one of them to actually make the real decision about where to locate the store. Employees who have some input into the decision making process are assumed to work harder gathering information. The manager and his team develop their own rules for their own internal decision making process, to encourage more effort and better information. If a subordinate actually makes the decision about locating the store, Aghion and Tirole would say that real authority differs from formal authority. The fact that the manager allows the subordinate to make the decision neither relieves the manager of the responsibility for the decision, nor does it imply that if push comes to shove and the manager and the subordinate disagree that upper management will side with the subordinate. If there is a conflict between the manager and his team, the firm's rule is enforced by the upper management: the manager decides. When a subordinate makes the decision it does not mean that the external default rule has been evaded, it simply means that real internal

⁹Despite the fact that papers in this branch of the organizational and contracting literature are about organizations of many people, the models have a tendency to focus on just two people. This makes it easier to write down and solve the models, it doesn't make them more insightful.

decisions are made in the shadow of the firms's rule. The external rules frame the relationship between the manager and his subordinates and their internal rules.

In their analytical model Aghion and Tirole show that "Real authority is determined by the structure of information, which in turn depends on the allocation of formal authority." (p. 1)

Value within this firm is produced by choosing the best locations for stores. How much information is produced in the work unit depends on the higher level default rules. Whether the prescriptive form of the firm's rule is followed in practice depends on the incentives the higher rule creates at the lower levels for individuals, both managers and subordinates. The firm's rules are "good" if they increase the value of the relationship (the information) produced by the manager and his subordinates, even if the actual decision maker is not the formal decision maker. The firm's rules *are not good because they are obeyed*, they are good rules if they increase the value of relationships within the firm. Table 3 depicts the firm's rule environment and organizational ecology. Note that it is not clear *ex ante* what the manager will decide to do in the work unit because Rule II allocates to him the authority to structure the work unit (just as the

carpenter has the authority in the laborer/carpenter pair).

The owners of the firm adopt secondary rules that give upper management the authority to make the rules that apply to the work units. Upper management has the option of making Rule

Rule I: Upper Management	Manager decides about locating new store: A prescriptive or default rule
Rule II: Manager	Manager could decide or
Rule III: Work Unit	Agreed upon rule within the work unit, including the manager about who decides.

I an explicit default rule, an explicit prescriptive rule, or leaving it vague. All three options affect whether the manager has the will choose to let the subordinates decide. The explicitly agreed upon formal rule may allow the manager to delegate the decision. That is fine. The upper managers in the firm may be trying to solve a more complicated problem with more dimensions and those problems may be better solved if the agreed upon rule is that the manager be strictly responsible (liable) for making the decision. In that case, the form of the rule would be prescriptive: the manager always decides because he is responsible.

Even if the higher level rules prescribe that the manager make the decision, however, the manager may implement internal rules within the work unit that allocate the decisions to subordinates. Since allocating the decision to subordinates is "against the rules" some may be tempted to argue that letting subordinates decide is a "norm" of behavior rather than an agreed

upon rule. But if every work unit in the firm makes up its own internal rules, how can a pattern of behavior be established to support a norm? What Aghion and Tirole show analytically, is that the incongruity between Rule I and Rule III may actually enhance the value of the firm. The rules in the rule environment do not have to match up with one another to add value. Congruent rules that are always enforced may not produce the best outcomes.

I hope you are beginning to see why the distinction between formal and informal rules gets us into trouble. The rules the individual laborer/carpenter pair adopt are no less “institutions” than the rules that the carpenters union and the laborers union adopt. The work unit rules are part of the rule environment of the firm. We should not be misled by what appear to be casual nature of “informal rules.” If the national level drug laws make possession and sales of marijuana illegal, and the state of Colorado and Washington make the possession and sale of marijuana legal, are the laws in Colorado and Washington “informal” because they are incongruent with the national laws?

Economists are fixated on contracts, for good reason. We have a tendency to contrast and oppose contracts and relationships. The laborer/carpenter relationship is an example of a relational contract ala Baker, Gibbons, and Murphy (2002):

Firms are riddled with relational contracts: informal agreements and unwritten codes of conduct that powerfully affect the behaviors of individuals within firms. There are often informal quid pro quos between coworkers, as well as unwritten understandings between bosses and subordinates about task-assignment, promotion, and termination decisions. Even ostensibly formal processes such as compensation, transfer pricing, internal auditing, and capital budgeting often cannot be understood without consideration of their associated informal agreements.

Business dealings are also riddled with relational contracts. Supply chains often involve long-run, hand-in-glove supplier relationships through which the parties reach accommodations when unforeseen or uncontracted-for events occur. (P. 39)

You can see how easy it can be to slip the “quid pro quo” and “unwritten understandings”

in to informal rules and so treat them as norms. But the terms themselves indicate that these are deliberate agreements. In their framework, there are formal contracts, written and agreed to, backed by third party (external) enforcement. Contracts, however, are inevitable incomplete. The agreements and relationships that exist within and between firms create the possibility that gaps in the formal contracts, situations which could not be anticipated or for which writing a formal contract would be too costly, can be bridged by the value of the relationship. Partners to a contract find themselves in circumstances not envisioned or covered by the contract, which might offer possibilities for strategic or expropriative behavior, yet both partners are willing to continue their relationship and their contract because of the value of their relationship. In this case, the gaps are in the contracts and the value of relationships gets the contracts over the gaps.

This way of thinking can get in the way of understanding how institutions actually work, however, because the reverse is also true: formal contracts can sustain relationships that have gaps in them. Lisa Bernstein studies how “private” legal systems in the cotton market support relational contracting in that industry. Cotton mills process raw cotton into thread and cloth. Cotton merchants purchase raw cotton from farmers and market it to cotton mills. Cotton is a highly variable commodity. Since mills do not want to maintain large inventories of raw cotton, merchants are often asked on short notice to provide cotton that meets narrow and complicated specifications. Mills and merchants in the United States are both located in a network of southern towns that have a long history in the cotton industry. Mills and merchants maintain deep and long lasting professional relationships that are strengthened by repeated interaction as well by social and family ties. It is not unusual for mills and merchants to maintain relationships through several generations, go to the same churches, and belong to the same clubs. Reputation

and repeated interaction matter in the cotton industry.

Contracts for cotton are often reached over the phone or by a handshake and later written up in a boilerplate contract. Detailed contractual terms are not negotiated in a written document, but through direct contact between the mill and merchant. Disputes do arise, however. Two trade associations have established rule based system to arbitrate disputes among their members. a combination of the American Cotton Shippers Association (ACSA) representing merchants and the American Textile Manufacturers Institute (ATMI) representing mills. In both cases the trade groups have established formal, agreed upon institutional rules and organizations to administer the rules that are binding on all the associations members. A requirement for membership in either association is that a firm agrees in case of dispute to submit to the binding rules of the association. The association rules are clear, formal, and applied narrowly and legalistically. The rules are impersonal, and the trade association goes to great lengths to insure that the rule are actually and perceived as enforced impersonally. The association rules are also default rules: the association only enforces the rules when disputes are brought to the association. The association does not patrol agreements between mills and merchants, the association rules are not prescriptively enforced.

Incongruence is an integral part of how the formal association rules are able to sustain higher value relationships between their members. Members are free to write incongruent provisions into their contracts that *cannot* be enforced through the association's procedures or in the regular court system.¹⁰ The agreements between merchants and mills are the real agreements that enable each pair of contractors to structure a better arrangement, even though terms of their

¹⁰Bernstein calls these "extralegal" provisions.

agreement cannot and will not be enforced by the third-party. Section II of the Bernstein's paper on the cotton trade is evocatively titled "The Theory of Legally Unenforceable Contracts." (p. 1767) This is precisely the sense in which the agreement between the laborer/carpenter pair is an unenforceable contract. There is no third party that will enforce an agreement that gives the labor the responsibility for driving nails, but that does not mean that such agreements do not exist or that laborer/carpenter relationships are not made more valuable by the existence of the nail rule that denies the laborer the authority to drive nails.¹¹

In the cotton industry, the market for raw cotton and exchanges between mills and merchants are governed by long term relationships. Bernstein shows how the contractual rules that will be enforced by the trade associations enable mills and merchants to bridge gaps in relationships. They sign boilerplate contracts that they do not enforce on each other and have recourse to the trade association rules, which are boilerplate rules, only in the rare case of a disagreement. This is why I am leery of using the word "hierarchy" to characterize the rule environment and organizational ecology. It simply isn't true that it is only the lower level relationships or the higher level contracts that bridge gaps in the other, conceptually both can bridge gaps in the other. It doesn't make sense, *ex ante*, to make one level "higher" than the other, except as an expository convenience. Similarly, it isn't clear that a "formal" contract, or agreed upon rule, should be higher in the rule environment than an "informal" relationship. That is why I suggest we jettison the terms formal and informal rules.

¹¹In another paper on the Uniform Commercial Code Bernstein argues that the Code is more valuable if the rules in the Code *do not conform* to actual business practices. Ayers and Gertler make the same point with respect to boilerplate default rules: they are more powerful coordinating tools if they do not correspond to usual business practices.

The third example comes from family law. In the 1960s and 1970s, American states began changing family law particularly the laws regarding divorce. Family law is a case of an institutional rule that most people think they know, but they really do not understand. Who reads the agreed upon rules governing marriage and divorce when they get married? In this critical area of life, governed by agreed upon institutional rules, most of us do not know what the rules actually are, yet we carry out our lives as if we do.

While the changes in 1960s and 1970s were unique to each state, in general they can be described as a move to “no fault” divorce.¹² The new institutional rules enabled a spouse to unilaterally dissolve a marriage, or for both spouses to bilaterally dissolve their marriage without having to prove that one spouse violated a rule that allowed the other spouse to get a divorce (like abuse or alcoholism). The new laws created a set of default rules that govern issues like child support, alimony, and property division. When a married couple comes to the court for a divorce, the court only applies these default rules only if the divorcing spouses cannot reach an agreement about these issues on their own. As Mnookin and Kornhauser (1979) document, less than ten percent of divorces actually involve a decision by a judge which settles disputes between the divorcing spouses and applies the default rule. While “hard” cases sometimes press the boundaries of the divorce laws, in most cases the judges apply well understood pre-existing rules for child support payments, alimony, and division of property.

Divorcing spouses must decide whether the default terms the judge will apply are better than the terms they can reach through arbitration and negotiation. Since the default rules the

¹²A brief introduction to the history of the changes can be found in Freed and Foster (1973) and the changes are tracked over time in a series of papers in the *Family Law Quarterly* by Freed and Foster and by Freed and Walker.

judge will apply are clear and impersonal, most divorce agreements are reached by the spouses (and their lawyers) and then rubber stamped by the court to be made official. Mnookin and Kornhauser emphasize how the existence of marriage default rules enhances the ability of divorcing spouses to use “private ordering” to determine how their marriage will terminate.¹³ It is only when private ordering breaks down, that the court applies the default rules. The divorce agreements are reached in the shadow of the law, but the divorce agreements are not bound by the law. These laws are truly default rules, not prescriptive rules. These institutional rules are not constraints.

Few people know what the laws governing divorce actually are when they get married, and many never find out unless they get divorced. Nonetheless, changes in marriage laws have a measurable and significant effect on behavior. Friedberg (1998) shows how the move to unilateral divorce raised divorce rates, at least in the short term. Stevenson and Wolfers (2006) demonstrate how the move to no fault divorce reduced the suicide rate among women within marriage. When one party could unilaterally end a marriage, finding that they were in a bad match, and ending the marriage earlier, reducing the amount of abuse and suicide caused by violence within marriage. Dee (2003), on the other hand, shows the violence by male spouses against female spouses did not change, but violence by females against male spouses increased by 21% when it became easier for either spouse to unilaterally exit the marriage.

No fault divorce laws are clearly default rules. The courts do not prescriptively enforce the rules, and will only make judgements when the parties cannot come to an agreement. It is

¹³As Mnookin and Kornhauser explain, and we should note, that divorce law in the United States before the no fault revolution was not structured as default rules in this way. In most societies, marriage law is not structured this way either.

implausible that many people knew the details of the divorce law and its changes, particularly teenagers whose behavior appears to have been affected by the changes. But it is quite easy to see how the effects of the changes in the agreed upon rules could have resulted in changing norms about marriage. Initially, people looking to end marriages don't find getting a divorce is easier than they expected, because they probably didn't know how hard it was before and so their expectations could not have changed. But a larger share of people contemplating or seeking divorce obtain one. Behavior patterns reflect that it is easier to get a divorce. Part of the response produces more marriages between people who are potentially at risk for bad outcomes. The number of marriages increases, including bad or abusive marriages, but so too does the divorce rate ending those marriages. The socio-economic outcomes associated with abusive marriages decline, particularly for women. Norms about marriage continue to evolve, even after the rules have stopped changing.

The only plausible channel through which these agreed upon rules affect behavior is through norms of behavior. Does that mean that the agreed upon rules do not matter, because people do not know what the rules are, and the behavior of people, both when they are married and when they are divorced, does not follow prescriptive forms? Now we are ready to go back to the question of institutions-as-rules and institutions-as-equilibria.

The institutions-as-equilibria approach does not look at rules as default rules. Although this section critiques the equilibrium approach, it is really a criticism of the explicit idea in North's definition of institutions that rules and norms can both be understood with the same underlying conceptual framework: a theory of institutions. The concept of rules in the

equilibrium approach is critiqued, not its concept of equilibria and norms as equilibrium phenomena in the social world. Norms of behavior are very important aspects of the social world, and I am in complete agreement with an approach to norms as constraints on behavior. On a meta-level, societies must reach arrangements that are self-enforcing in the sense that everyone has incentives to abide by the arrangements because, at the level of societies, there can be no external third party enforcers. That said, if we want to think about societies as equilibria, then we have to acknowledge that agreed upon rules are part of the social arrangements necessary to make larger societies possible. Further, that societies with agreed upon identity rules are much more fragile than societies with agreed upon impersonal rules. We can't wrangle that insight out of the institutions-as-equilibria's approach to rules and enforcement.

The hinge of the argument opens and closes on the spurious distinction between formal and informal rules. Confusion arises because, literally, norms of behavior become informal *rules* in North's formulation. In the nail rule example the agreement between the laborer/carpenter pair creates a rule. The only way, however, that we as outside observers can determine what the rule is is to ask them, or to observe how they behave. Unfortunately, behavior under a default rule may not produce behavior that enables us to infer what the rule is from behavior. As a result, in situation where rules exist we may not be able to see them in operation. Another example may help here.

Families are a common type of organizations in all societies. They are structured in very different ways in different societies, but virtually all families create some internal division of labor. My family is no different. My wife and I have agreements about how we divide up tasks. One of our agreements is that I take out the garbage and my wife does the dishes, another is that

my wife pays the bills and I balance the checkbook (we call it “doing the bank.”) These agreements produce rules that help order our relationship, but they produce different patterns of behavior in relation to the agreement. A half full garbage can bothers my wife much more than it bothers me, so my wife often takes out the garbage. If you stood outside our house and counted the times that each of us took out the garbage, you would infer that our rule allocated responsibility for taking out the garbage to my wife sometimes, me other times, and probably would deduce that we have no rule for taking out the garbage. On the other hand, my wife always pays the bills and I always balance the checkbook.

In either case when my wife asks me to “take out the garbage” or “do the bank” I try to do it, quickly if possible. The behavioral pattern in our family is that my wife always pays the bills, we share taking out the garbage, and I always balance the checkbook. The behavioral patterns do not correspond perfectly to the agreed upon rules. The garbage rule is a default rule. Unlike the nail rule, it is a rule that applies solely within our family, there is no external third party like the unions to enforce it. My wife has the option of asking me to take out the garbage, but she does not have to. If we cannot agree on who should take out the garbage, I take it out, that is the default. An external observer could never figure out what our garbage rule actually is, while, in contrast, the bank rule is immediately obvious in our behavior because my wife never balances the checkbook. An observer could observe the bank rule, but not the garbage.

The laborer and carpenter may reach an agreement about driving nails that produces behavior an external observer cannot classify as following a rule. There may be no discernible pattern to who drives nails, but that does not mean that the laborer/carpenter pair have not reached an agreement about a rule that structures their relationship. In that sense, their

agreement is an equilibrium. But it is also true that their agreement is a unique outcome of their relationship and it *is a deliberate agreement*.

Both the nail rule and the garbage rule would be classified as informal rules. Because it is difficult, perhaps impossible, to infer the existence of the rule from behavior patterns, we are likely to attribute the way nails are driven and garbage is taken out to something, but what do we call that something? Should we attribute the behavior to norms because we cannot directly link the behavior to a observable rule?

The key insight, and it a profound insight, of the institutions-as-equilibria approach is that patterns of behavior in the external world create expectations about how the world will behave and how the world will react to my behavior. Those expectations set up costs and benefits that enter into my decision process.

The crucial point is that while a “rule” may serve as a coordination device, it is fundamentally the expected behavior of others, rather than the rule itself, which motivates people’s behavior. A similar logic can be used to examine economic, political, and social institutions even in situations involving specialized actors and more complex formal “rules”. From the institutions-as-equilibria perspective, it is always ultimately expectations about the behavior of the other actors (including those in specialized enforcement roles such as police, judges, etc.) that create the institutional constraints which mold people’s behavior, and all such behavior must therefore ultimately be explainable endogenously as part of the equilibrium. (Greif and Kingston, pp. 14-5.)

The institutions-as-equilibria critique of institutions-as-rules revolves around why people follow rules. In a simple Hobbesian or neo-classical model of rule following, people follow rules because the penalties that will be imposed on them by the rule enforcers create costs that are less than the benefit of breaking the rule. That simple model of why people follow rules underlies the institutions-as-rules approach:

Within the institutions-as-rules view, the enforcement of the rules is considered as a distinct issue from the formation and content of the rules themselves. Enforcing the rules

involves “enforcement costs”. The formal and informal rules, together with their “enforcement characteristics” constitute the institutional structure within which interactions occur. Thus, the institutions-as-rules approach employs a rational choice perspective to study the formation of institutions, but a theory of motivation – explaining why people follow particular rules of behavior – is not integrated into the analysis. (Grief and Kingston, p. 14)

Both the institutions-as-equilibria and institutions-as-rules assume that rules matter because they coordinate the relationship of individuals, of social behavior, because when the rules are followed they create expectations of behavior. As the nail rule and garbage rule examples show, as the logic of Aghion and Tirole shows, and the empirical analysis of Bernstein, of Mnookin and Kornhauser, of Freidburg, of Stephenson and Wolfers, and of Dee show, default rules can coordinate behavior even if they are not followed. If they are not followed how should we conceptualize “enforcement” of those rules? If a default rules cannot be observed in the behavior is is supposed to govern how can the behavior it induces create a norm? Or should we conceptualize the enforcement of these default rules within the reality that in most, probably all societies, there are rule environments and organizational ecologies in which many rules are incongruent and therefore *not all of the rules can be enforced*.

When North defined institutions “more formally are the humanly devised constraints that structure human interaction” he placed rules and norms both in the “rules of the game” because the existence of both norms and rules could constrain individual choice. Norms ala the institutions-as-equilibria approach *are always followed*. Norms only exist if a pattern of behavior creates expectations in individuals and those expectation serve as a constraint on their choices. Norms appear at all scales of society, just like rules. They appear in families, neighborhoods, organizations, schools, churches, regions, states, and nations. No one worries that the norm “environment” contains norms that are in conflict with one another, as they clearly

are at every level of every society. Because norms are always followed, but not agreed to, there is no incongruity in the existence of conflicting norms. Agreed upon rules however, are always agreed to but not always followed.

Why have we been insisting that rules and norms are basically the same social phenomena that can and should be explained by the same theoretical framework? Because we, and certainly North, take the “always followed” part of norms and applied it to agreed upon rules. We assume that rules matter because they are followed. Asking why people follow or don’t follow default rules is problematic in a way that asking why people follow or don’t follow norms is not. Why people don’t follow norms seems obvious, people have different experiences and beliefs about the world. But how could agreed upon rules be coordinating devices if people don’t follow them?

This isn’t just North’s fault. We who study institutions have developed fallen into way of thinking, a pattern of behavior, in which the subject matter of institutions includes both agreed upon rules and norms of behavior, and we have both searched for and imposed the idea that the same underlying social process works the same for both norms and rules. Perhaps it is the fact that many, indeed perhaps most of the agreed upon rules that apply to us are not rules that we “agreed” to, in the sense of active agreement and active consent. We passively accept the existence of rules in the same way that we must passively accept the existence of norms, even if we do not like the rules or the norms. But that does not mean as social scientists we should identify rules and norms as the same social process.

Rules are agreed to in a way that norms simply are not, and this gets us back to the problematic nature of the formal/informal rule distinction. All rules are agreed to, but the

process by which rules are agreed to varies widely from society to society, and even within a single society the process of rule agreement varies depending on where the rule is agreed to in the rule environment and organizational ecology. So rules that we have trouble identifying, rules that are agreed to under secondary rules that are not immediately obvious, get pushed into the informal rule category.

We classify informal rules as informal because if we cannot see the secondary rules by which they were agreed to, then we can't really explain them, and there is no good reason not to treat informal rules in the same category as norms. This, despite the fact that a defining feature of norms is that they are not agreed to. That norms are not agreed to but still constrain individual choice is, indeed, a fundamental insight of the institutions-as-equilibria approach.

This suggests that a fundamental distinction between rules and norms is that rules are agreed to and norms are not. Although we often focus on enforcement of agreed upon rules, enforcement is problematic once we come to grips with incongruent rules. Norms are always enforced within the relevant group. That powerful insight is always conditioned by the fact that enforcement is internal to the individual who perceives the pattern. It cannot rule out, for example, that people are influenced by patterns they think they see but that do not exist. The close association of norms with beliefs lays here. Beliefs about rules also matter. There is no doubt that beliefs are fundamental to explaining individual choice, which is why North's later work veered more to focus on cognition and beliefs in *Understanding the Process of Economic Change*.

Asking why people follow rules is not the same question as asking why people follow norms. Agreed upon rules are often not followed. A default rule is only enforced in specific

circumstances. Outside of those circumstances individuals need not follow the rule, although their behavior may conform to the rule, and there is no enforcement of the rule. That does not, however, mean that the rule does not affect behavior. It can. Enforcement is not critical to a rule's effectiveness at altering behavior. Although as noted in the nail rule example, how a default rule is enforced when circumstances call for enforcement do affect how the rule works. Identity enforcement of a default rule provides much less coordination than impersonal enforcement.

People who work on institutions, theoretically and empirically, often find the equation of norms, rules, values and beliefs with institutions troubling and a cause for concern.¹⁴ In their paper on “Institutions and Culture” Alessina and Giuliano give voice to the concern:

Semantically speaking, we find it counterproductive and confusing to label culture (meaning values and beliefs) as informal institutions. We find it confusing to label “everything”—from, say, the level of reciprocal trust in a society to constitutional rules about voting systems—as institutions. Clearly—this is the crux of our paper— culture (or informal institutions) and formal institutions are interrelated, but the label “informal institutions” implies that formal institutions determine informal ones, and that the latter are of secondary importance. Once we agree that formal and informal institutions interact, and that either one may cause the other, then identifying certain values and beliefs as culture or informal institutions becomes merely a matter of semantics. We prefer the term *culture* over *informal institutions*; we find it more appropriate and less confusing. Similarly, for brevity, we sometimes refer to formal institutions simply as *institutions*.” (Alessina and Giuliano, p. 902)

Alessina and Giuliano are on the right track, institutions and culture should be considered separately and we do want to understand how they interact. But the formal/informal distinction intrudes. There are many agreed upon rules that we would classify as “informal” and should not

¹⁴When Barry Weingast and I were writing *Violence and Social Orders* with Doug North we had many conversations about beliefs and values (see chapter 1 of that book), but Barry and I always wondered where exactly they fit into institutions. We did not wonder whether they were important, they are, but where they fit into a theory of institutions wasn't clear.

be considered part of culture (although they are clearly an integral part of any society). This, however, is not just a question of semantics. It is a fundamental question of the elements we need to use to explain how the world around us works. It is about the ontology of the world, as the philosophers would call it. What pieces of the world are we trying explain? How they interact will depend on what the pieces are. Agreed upon rules and norms should be different pieces, they should not both be part of “institutions,” even institutions broadly defined.

In table 2 depicting the nail rule environment and organizational ecology, where we should draw the line between formal and informal rules is not at all clear. Alessina and Giuliano approach seems to suggest that we should draw it below the State Safety Commission and above the unions. Then everything below that break point becomes a “informal rule” and therefore part of culture rather than institutions. If we draw a horizontal line in the table and say above the line the rules are formal and below the line the rules are informal, then it seems inevitable that we will fundamentally mischaracterize how agreed upon rules work. If we insist on putting rules into a hierarchy, and the formal/informal distinction is hierarchical, we will miss this critical element in how rules really work. As the trade association example shows, either or higher and lower orders of organization may support each other; the relationships and contracts that interact can be at both levels.

We surely do not want to associate formal rules only with governments. Here is how Alessina and Giuliano on one hand, and Greif and Kingston on the other characterize North: “In North’s theory, formal rules are created by the polity, whereas informal norms are ‘part of the heritage that we call culture.’ Institutions, he says, are ‘the rules of the game.’” (Alessina and Giuliano, pp. 901-2). “The institutions-as-rules approach, following North (1990, p. 3),

identifies institutions as ‘the rules of the game in a society’, including both “formal” rules such as constitutions and laws enforced by the state, and ‘informal’ constraints such as ‘codes of conduct, norms of behavior, and conventions’, which are generally enforced by the members of the relevant group (North, 1990, p. 36).” (Greif and Kingston, p. 14) If we read those sentences literally, and I think the quotes give too narrow a reading of both papers, then everything in Table 2 below the level of the State Safety Commission would be an informal rule. Even if the quotes are too narrow, where do we draw the line? The union default rules would get swept into culture by Alessina and Giuliano’s framework. If the rules were not followed they would be irrelevant in Greif and Kingston’s framework. That is not an outcome that anyone wants.

Where do we go from here? My thinking about institutions, rules, and norms changed as I wrestled with explaining how and why a few societies, beginning with Britain and the United States between 1830 and 1860, were able to create and enforce impersonal rules: rules that treated a wide swath of residents or citizens the same. Impersonal rules are never universally impersonal. Impersonal rules were usually called “general laws” in the United States and in the 1840s and 1850s when states began modifying their constitutions to require general laws for incorporation and a wide range of other activities, they did not intend the rules to apply to women or slaves. They were hardly universally impersonal, but the rules applied to a wide enough portion of the American public that it became difficult, and in some cases impossible, for the political process to craft rules and enforcement that whose form could differ across individuals with different social identities, no matter how identity was defined.

Although the constitutional provisions and the laws that they resulted in were all couched

in prescriptive terms, many of the new general laws were default rules. General incorporation laws were critically important in opening entry and access in the early 19th century America economy, but states *never* actively checked to make sure corporations followed their charters. These were never prescriptive rules. You cannot, however, discern from reading the statutes, you need to know the history. Legal scholars, theorists, historians, and applications all recognize that there are many rules and laws that are not actively enforced, but the insight has not built into anything more substantial than categorization. The social sciences have, by and large, been content to use a Hobbesian conception of rules, whether the rules are public and private, and focus on the incentives needed to enforce the rules. Implicitly, and sometimes explicitly, we assume that the purpose of the rules is to assure that behavior conforms to the rules. The institutions-as-equilibria approach chastens the institutions-as-rules approach for not have a theory of enforcement. “Thus, the institutions-as rules approach employs a rational choice perspective to study the formation of institutions, but a theory of motivation – explaining why people follow particular rules of behavior – is not integrated into the analysis.” (Greif and Kingston, p. 14)

We should be chastising both the institutions-as-rules and institutions-as-equilibria approaches for having a deficient theory of rules, what they are, and how they work. This paper has sketched out such a theory. It must include some concept of the rule environment and organizational ecology and an awareness of how rules in different organizations interact. Not just a hierarchy of rules, but a rule environment where all the rules and organizations potentially affect each other. It must include prescriptive and default rules as two primary species of rules and recognize that enforcement of the two types of rules is fundamentally different. It must

recognize that rule environments are made up of primary and secondary rules, and not only understand how the two interact but that in many circumstances a particular rule is regarded as a primary and secondary rule simultaneously in different parts of the rule environment and organizational ecology. Finally, it must appreciate that identity and impersonal rules differ in both form and enforcement, and while there were certainly impersonal rules that applied to small and most likely privileged groups before the 19th century, for most of recorded human history, agreed upon rules have been identity rules.

That leaves us with the last question: what to do with the term “institution?” My own inclination is to recognize that the study of institutions in the social sciences, law, history, and philosophy has its own long history, and that the subject matter has been highly varied and defined in a multitude of ways. But then to argue that the time has come to distinguish between institutions as agreed upon rules, and other types of social phenomena, like norms, beliefs, and values that interact closely with institutions, but are better delineated by a different conceptual framework than institutions. Alessina and Giulianos contrast between institutions and culture is a step in the right direction, but one that does not pay enough attention to the nature of rules.

I strongly suggest that we abandon the terms formal and informal institutions as categories of analysis. It is a way to organize one’s thinking, but not in the end a way that leads to better ideas. Instead, institutions are agreed upon rules. Institutions exist at every level of society, from families to corporations and organizations, to nation states, and even to international organizations, all of which agree upon rules for their internal relationships and their relationships with other organizations. Deliberate agreement is what makes a rule an institutional rule. As John Commons said in 1931 “An institution is defined as collective action

in control, liberation and expansion of individual action. Its forms are unorganized custom and organized going concerns.” He was right about collective action. Reaching an agreement about what rules are is the essential element of institutions. Rules are deliberate attempts to structure human relationships. Most institutions either don’t work at all or don’t work as their creators envisioned, but that shouldn’t deter us from studying them. The world has changed over the last ten thousand years and even more dramatically in the last 200 years because of the institutions that human societies have collectively agreed upon. Commons wasn’t wrong about the importance of “unorganized custom,” but he should have included it in culture and focused on collective action: deliberate attempts to agree on rules that will govern human interaction.

Incomplete References:

- Aghion, Phillippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105 (1), pp. 1-29.
- Alessina, Alberto and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53(4), 898–944.
- Ayres, Ian and Robert Gertner. 1989. "Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules." *The Yale Law Journal*. Vol 99, no 1, pp. 87-130.
- . 1992. "Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules." *The Yale Law Journal*, Vol. 101, No. 4 (Jan., 1992), pp. 729-773.
- Baker, George, Robert Gibbons and Kevin J. Murphy. (2002). "Relational Contracts and the Theory of the Firm." *The Quarterly Journal of Economics*, Vol. 117, No. 1 (Feb.), pp. 39-84.
- Baker, George, Robert Gibbons and Kevin J. Murphy. (1994). "Subjective Performance Measures in Optimal Incentive Contracts." *The Quarterly Journal of Economics*, Vol. 109, No. 4 (Nov.), pp. 1125-1156
- Bernstein, Lisa E. 1996. "Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms." *University of Pennsylvania Law Review*, 144 , pp. 1765 - 1821.
- Bernstein, Lisa E. 2001. "Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions," *Michigan Law Review*, 99, pp. 1724-1790.
- Bernstein, Lisa E. 2015. "Beyond Relational Contracts: Social Capital and Network Governance in Procurement Contracts," *Journal of Legal Analysis* 7, pp. 561- 621.
- Gibbons, Robert. 2010. "Transaction-Cost Economics: Past, Present, and Future?" *The Scandinavian Journal of Economics*, Vol. 112, No. 2 (June), pp. 263-288.
- Greif, Avner and Christopher Kingston. 2011. "Institutions: Rules or Equilibria?" in N. Schofield and G. Caballero (eds.), *Political Economy of Institutions, Democracy and Voting*. Berlin: Springer-Verlag.
- Greif, Avner. 2006. *Institutions and the Path to the Modern Economy*. New York: Cambridge University Press.
- Hennessey, Jessica and John Joseph Wallis. 2017. "Corporations and Organizations in the

United States after 1840.”, in Lamoreaux and Novak, ed. *Corporations and American Democracy*, Cambridge University Press, 74-105.

Hindriks, Frank, and Francesco Guala. 2014. “Institutions, rules, and equilibria: a unified theory.” *Journal of Institutional Economics*, 11: 3, 459-480.

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