Customary rules versus designed rules

Karol B. Zdybel[1]

European Doctorate in Law and Economics, Institute of Law and Economics, University of Hamburg

Abstract:

This paper compares two stylized mechanisms for establishing social rules: custom and deliberate design. Customary rules emerge based on past experiences in similar social interactions, as seen in customary international law or archaic laws. In contrast, rule design involves a rule-maker, which is typical of modern-day legislation or great historical codifications. Unlike most earlier works on similar topics, we consider an incomplete information scenario: our approach realistically assumes that agents have limited knowledge of each other's preferences and objectives. This limitation is shown to be consequential for rules developed through both mechanisms. We find that customary rules tend to be less internally differentiated because easily replicable and observable practices are more likely to become normatively expected than more case-specific ones. When efficiency requires high complexity, designed rules may have an advantage. Moreover, the participatory nature of custom formation makes it more responsive to underlying economic forces but unresponsive to innovations that bring about Kaldor-Hicks efficiency improvements. Conversely, Kaldor-Hicks improvements are compatible with deliberate rule design. Because designed rules effectively serve as third-party coordination devices, their specific components cannot be vetoed. However, the indivisibility of the incentives to follow designed rules is double-edged: while potentially enabling higher degrees of efficiency, it simultaneously makes deliberate rule design susceptible to manipulation by vested interests and to errors resulting in adverse unintended consequences.

Keywords: Customary law, custom, legislation, rule design, institutions

JEL-classification: B52, C73, D83, K00, K10

---

[1] ORCID: https://orcid.org/0000-0002-3432-6311

The author thanks Michael Faure and Stefan Voigt for their helpful comments on an early version of this draft.

0. Introduction

How can social rules be developed? Despite the broad nature of this question, it has consistently drawn the attention of scholars across various disciplines – evidence of its longevity and ongoing importance. This paper revisits this familiar issue but adds a significant and often omitted factor: incomplete information about objectives, preferences, and needs of other actors. The paper begins from a commonsense premise: in realistic scenarios, agents often lack full knowledge of decision-relevant circumstances. For instance, a tradesman might repeatedly face a dilemma between improving product quality, which could delay delivery, and meeting the delivery deadline at the expense of quality. The tradesman's decision on whether to delay delivery depends on the reaction he anticipates from his customers, which in turn depends on their private preferences – whether they prioritize quality or punctuality. Similarly, during a major epidemic, individuals may limit social activities to reduce transmission, expecting each other to disengage from less urgent ventures. Since the value of various activities differs among individuals and these valuations often remain private, the question arises: what kind of rules can we expect to develop?

Building on the backdrop of incomplete information, we compare two stylized mechanisms for establishing social rules: custom and deliberate design. These two represent the broader categories of unintentional and intentional institution-making. Customary rules develop when agents rely on past experiences to form expectations about others' actions and adjust their own accordingly. In contrast, designed rules are supplied by a third party, with expectations explicitly set in advance. This paper systematically examines how these two mechanisms for creating rules compare in light of four criteria: rule complexity, rule efficiency and bias, susceptibility to manipulation by vested interests, and susceptibility to error.

Among these criteria, rule complexity means the internal differentiation within a rule, or the number of distinctions it makes between individual cases; more distinctions imply greater complexity. Capacity for efficiency and inefficiency refers to how efficient or inefficient a rule can be, how its efficiency is affected by changes in the underlying economic realities, and how it can exhibit a bias that favors one party instead of another. Manipulation denotes a situation when the content of a rule is influenced to systematically advance the vested interests of a particular party at the expense of the general economic efficiency of the rule. Error refers to the presence of adverse unintended consequences of a rule.

While there are numerous studies on customary or "spontaneous" institutions (see, e.g., Powell and Stringham, 2009), relatively few have conducted comparative analyses between customary and deliberate institution-making from a law and economics perspective. One notable exception is Parisi (2001), who

compares legislation and custom as two "institutional designs of rulemaking." Parisi identifies legislation with political procedures that can be analyzed through public choice theory, suggesting that democratic lawmaking is burdened with numerous agency problems and adverse incentives, which are likely to render the resulting rules inefficient. Simultaneously, Parisi argues that custom suffers from the collective action problem: agents may face inadequate incentives to create socially efficient customary rules. Rossi and Spagano (2018) attempt to explain the codification of customary rules used among early modern merchants. They argue that the cost of disseminating rules played a major role in this process. Because it is more difficult to familiarize oneself with customary practices in larger social networks, written text began to enjoy an advantage as the relevant markets expanded.

We believe that by seriously considering incomplete information, the comparison developed in this paper can improve upon previous studies. Earlier contributions often focused on the stability and efficiency of rules that are fully developed and complete. In contrast, this study highlights the importance of the rule-creation *process*. Typical models portray rules as equilibrium-sustaining expectations about the behavior of other parties. If travelers expect each other to keep on the right side of the road, these expectations are transformed into a pattern of behavior that consistently confirms these expectations. Importantly, agents form their expectations with full knowledge of each other's preferences and objectives, understanding the outcomes desirable to others and how they are related to their own desirable outcomes. In other words, agents possess full knowledge of the game form, which implies the knowledge of the possible equilibria. Since there is usually more than one equilibrium, institutional analysis must address the issue of equilibrium selection: which predefined equilibrium is most likely to materialize in reality? (Mahoney and Sanchirico, 2001; Knight, 1992).[2]

[2] Aoki (2001) suggests that there are two principal methods for analyzing social rules or economic institutions using game theory. One approach uses variants and refinements of the subgame-perfect solution concept, according to which agents choose what they believe to be the optimal course of action at each decision point. The other approach represents institutions as fixed strategies employed by different segments of the agent population. These strategies may differ in efficiency and in how well they perform when clashed with other strategies – two factors co-determining the rate at which each strategy is reproduced in the population. Over time, certain strategies may come to dominate the agent ecology. The subgame-perfect equilibrium and related concepts assume that agents are fully rational, but they typically rely on the assumption of complete information. On the other hand, evolutionary models allow for some flexibility regarding information but treat agents as following a predetermined program, making them less suitable for representing rational agents similar to human beings. In contrast, this paper retains subgame perfection (and thus strict rationality) while introducing the more realistic assumption of incomplete information.

The logic of selecting between predefined, known equilibria is summarized in the following statement: "If all the actors can focus on something that distinguishes one of the two equilibria, then over time they will be able to establish a regularity in behavior that, when confronted with similar interactions in the future, can serve as a guide to coordinated action. (…) [A]ll of this depends on the actors' initial ability to focus on a salient difference between the two coordination equilibria." (Knight, 1992:100) In other words, the possible ways to govern a social interaction are assumed to be revealed by default; the problem of establishing institutions is reduced to selecting one of these methods. While this approach is popular and undoubtedly useful, it overlooks an important point: agents not only need to select an equilibrium to coordinate on, but also must first discover which equilibria are even possible to play.

Therefore, unlike the logic founded on the assumption of complete information, this paper considers a realistic scenario where agents have limited knowledge about outcomes others consider desirable or undesirable. This approach aligns with the observation by Hayek (1945) that socially useful knowledge is typically dispersed in society among heterogeneous individuals. Consequently, effectively utilizing this knowledge – i.e., incentivizing its disclosure and transmission in society – poses a challenge in its own right. This paper claims that processes through which institutions are established face the same challenge, and these challenges may affect the content and properties of the resulting institutions. The process of discovering equilibria may be consequential for the shape of social rules the long run.

To advance its case, the paper employs a game-theoretical framework. Specifically, it uses a model of a repeated game with incomplete information to account for less-than-perfect knowledge. However, given that the argument may be sometimes technical, an informal summary is provided first.

In essence, we propose that customary mechanisms for rule creation resemble finitely revealing equilibria known from game theory. In this type of equilibrium play, agents share private information in a limited number of steps. This sharing may involve actions such as making a cooperative effort for the first time, punishing behaviors when the offending party does not expect punishment, taking risks before others follow suit; it can also take the form of forgoing actions, withholding cooperation, or deliberately ignoring transgressions. These actions lead other agents to update their beliefs about the private, decision-relevant characteristics of others, which then shapes their expectations about their likely future actions. In other words, agents use trial-and-error methods to identify viable ways of long-term cooperation. After this formative phase, where beliefs and expectations are updated, customary rules become firmly established, resulting in a

long-term stable equilibrium in which players' beliefs and expectations no longer change.

Generally, the formation of more complex customary rules requires more initial steps, or more investment in the acquisition and interpretation of relevant information in the formative phase. We claim that simpler customary rules have the advantage of offsetting this requirement and therefore are more likely to develop. Because they are cheaper to communicate, observe, and interpret, simple rules have a higher chance of being widely adopted and followed. Simultaneously, we suggest that the participatory nature of custom formation tends to produce Pareto improvements in the formative phase, i.e., rule content evolution that gradually improves the utility position of each agent. It also makes it relatively more responsive to underlying economic forces, such as the cost-benefit ratio of various activities, and harder for vested interests to capture.

In turn, establishing rules through design – such as legislation or codification – is likened to constructing correlated equilibria. A correlated equilibrium refers to a situation when agents use a third-party coordination device to structure their interactions. In this scenario, strategies used in the long-term equilibrium are not "discovered" by the agents but deliberately preconfigured. As long as agents expect each other to follow these unique preconfigured strategies, none has an incentive to defect. Correlated equilibria offer an important informational advantage: agents can observe whether a third-party device induces a correlated equilibrium in a number of steps that is independent of rule complexity. This means that designing rules can generally allow for greater complexity and sophistication. However, greater complexity does not necessarily lead to more efficient rules. The incentives to rely on third-party coordination are indivisible, meaning that agents cannot selectively reject its parts without undermining the entire process. Consequently, deliberately designed rules can accommodate significant inefficiency, whether resulting from manipulation by vested interests, design errors, or other sources.

The remainder of the paper explicates the argument outlined above in more detail. It is structured as follows: Section 1 elaborates on the alternatives between custom and institutional design from philosophical and historical perspectives. Section 2 introduces the formal model. Sections 3 and 4 use this model to explore custom and design as two distinct mechanisms for establishing social rules. These sections examine conditions critical for the functioning of both mechanisms and conducive to the qualitative differences between the two. Section 5 discusses the findings, taking into account the comparative criteria specified above. Finally, Section 6 concludes.

1. Philosophical and historical background

Conventional wisdom suggests an obvious mechanism for establishing social rules: design. A neutral third party or the interested parties themselves may create and announce commonly known standards of behavior. For example, how should a rancher compensate a farmer when the former's cattle destroys the latter's crops? An obvious answer is that a lawmaker can proclaim rules that specify the compensation regime. Notably, designed rules are subject to rational control: they are intentionally constructed to govern future interactions between individuals and can be directed to serve the purposes envisioned by the rulemakers.

On the other hand, a significant strand of scholarship downplays the importance of rule design. This perspective proposes that functional social order may emerge without recourse to predefined, rationally construed rulesets. Instead, it emphasizes the prevalence and efficiency of customary rules (e.g., Hayek, 2013 [1973]; Sugden, 1989; Taylor, 1982). In this context, custom can be understood broadly to represent rules derived from past experiences in the same social dilemma. So-defined custom encompasses a variety of traditions, social norms, or historically established conventions. Scholars tend to agree that rule production mechanisms based on custom are especially common in areas like international law, social orders of pre-literate men (i.e., primitive law), and everyday social norms (see, e.g., Barkun, 1968; Guzman, 2008; Young, 2015).

The tension between these two stylized models of rule emergence has been observed by legal scholars and legal philosophers. For instance, Raz (1994:371) describes an institution-oriented model of the rule of law, which "requires elaborate bureaucratic machinery with meticulously observed and policed procedures, [...] and which require for their success anonymous impartial institutions, inhabited by impartial strangers." This model assumes that publicly established, prospective, and general rules are necessary for its implementation. Raz contrasts institution-oriented models with models of the rule of law that treat legal rules as commonly shared traditions.

Likewise, in't Veld (2023:265-266) distinguishes between the bureaucratic and tradition-oriented ideals of the rule of law. Crucial to the bureaucratic ideal is its deliberate future orientation; it requires that "[t]he law should be publicly laid down so that people are able to plan their lives accordingly." In turn, the tradition-oriented ideal can be conceptualized as "a set of practices which evolved over time and withstood the test of time." This model views the rule of law as an organic accumulation of customs and practices that are legitimized through their longevity and practical efficacy, rather than through formal rule-making processes. In a related vein, Hadfield (2017) describes social control in preliterate societies. She contrasts tacit rules that emerge from long-standing traditions, which are organic

and informally upheld by community consensus, with those that arise from deliberate legal orders that are supported by sophisticated legal infrastructure, including professional lawmakers.

A similar tension is present in Hayek's theory of spontaneous order and its antitype, "made" order (Hayek, 2013). While Hayek's concept of "order" is broad, it can be concretely applied to social rules. According to Hayek, rules may arise as products of unsupervised evolutionary processes. Through the interactions of individuals each pursuing their own interests, social rules can emerge without being purposefully constructed. This spontaneous emergence results from the gradual convergence of agents' actions and beliefs into a consistent pattern of group behavior. Conversely, rules can also be deliberate products of rule-making activities. This rule creation mechanism involves intentional design by third parties, aimed at promoting specific social objectives. In short, spontaneous orders emerge when third parties do not intentionally design rules; "made" social orders are characterized by third-party entanglement with a specific purpose in mind (Luban, 2020).[3] Further examples of the contradistinction between broadly conceived customary and deliberate methods of establishing institutions can be found in institutional economics (Knight, 1992) and legal theory (Parisi, 1995; Hart, 1961).

Historical evidence suggests that customary and deliberately designed rules often played complementary roles in regulating social behavior. Both rule-making mechanisms were used to address the same social dilemmas, often in immediate succession. For example, consider the prevention and repression of various forms of sexual misconduct in antiquity. On legislative grounds, it might seem that pre-imperial Rome was lenient toward sexual offenses. During the Roman Kingdom (753-509 BC) and most of the Roman Republic (509-27 BC), rape was not criminalized (Gardner, 2008). Moreover, it is uncertain whether any civil remedy designed specifically for rape was provided in republican legislation. Most likely, rape was not treated as a standalone civil wrong. Depending on the circumstances, it could be addressed under categories such as assault or abduction, for which there were established legal remedies in written law (Lintott, 1968; Gardner, 2008). In short, protection against sexual offenses appears to have been largely absent from Roman law until the late Republic.

[3] In a way corresponding to the Hayek's argument, Leoni (1993:7) considers custom a viable rule-making method for a society: "fewer and fewer people now seem to realize that just as language and fashion are the products of the convergence of spontaneous actions and decisions on the part of a vast number of individuals, so the law too can, in theory, just as well be a product of a similar convergence in other fields." He insists it is not only a viable but a desirable method: "The law-making process ought to be reformed by making it mainly, if not only, a spontaneous process, like that of trading or of speaking." (Leoni; 1993:134)

The absence of explicit legal statutes against rape suggests a reliance on alternative, non-codified, societal mechanisms to manage and mitigate issues related to sexual misconduct. Indeed, unwritten customary rules pertaining to unwelcome sexual behavior seem to have been robust, rigorously upheld, and often enforced through severe and violent means when transgressions occurred. Private initiative played a major role in preventing and punishing sexual misconduct. As Lintott (1968:26) claims, "the crimes of rape and adultery were (…) the subject of private revenge throughout the Republic." While such revenge was not explicitly recognized in legislation, it was considered a legitimate response by the general public and thus played a role in preventing and repressing offenses of sexual nature.[4]

When legislation finally interfered, it primarily aimed to legitimize preexisting customary rules while occasionally delegitimizing others. This pattern can be seen in legislation addressing sexual harassment. Sexual harassment becomes recognized as a private wrong in the late Roman Republic; remedies are granted against following or soliciting women and, most interestingly, against drawing a female's guardian away. This legislative intervention appears to have been initially driven by a desire to reinforce existing social norms regarding female chastity, particularly the expectation that women should be accompanied by a male (Gardner, 2008). In turn, when legislation against various forms of vice, including rape, adultery, and other sexual misconduct is enacted in the early Empire, previously strong social norms of sexual conduct seem to have largely vanished or become significantly relaxed, as evidenced by the preoccupation of the legislation with concubinates (Nguyen, 2006; Gardner, 2008).

The substitution of designed rules for preexisting customs was often done explicitly through the acts of writing down (black-lettering) of previously prevalent customary rules. For example, in the era of Greek colonization, customary laws became "codified and written down with the addition of new laws framed by the lawgiver to correspond to the exigencies of his time" (Smith, 1922:187). Likewise, the Law of XII Tables, traditionally considered the first piece of Roman legislation, seems to have been "nothing more than a codification of such [customary] law, with perhaps a few legislative innovations" (Schiller, 1938:275; Westbrook, 1988).

The same was true of orally transmitted customs in Medieval Europe. Since the 12th century, written inventories of customs specific to a given place (e.g., town

[4] The extensive reliance on self-help seems to have been one of the general features of the social order of the early Roman Republic. Norms regulating the legitimate use of self-help have been gradually incorporated into written law. According to Lintott (1968:34), "Roman legal procedure was originally modeled on ritualized self-help, and for its successful functioning it relied on self-help. Private action was its foundation, and so it cannot be surprising that it allowed the individual so much scope to right his own wrongs."

or province) have begun to surface. They originated as mere attempts to provide a tangible list of rules binding within a given community and initially were not treated as authoritative sources of law. Writing down customary rules originally happened in an unofficial, private manner. However, such private inventories are soon replaced by *official* statements of customary rules authorized by appropriate authorities; eventually, they become the only valid sources of rules, replacing living customs altogether (Glenn, 1997; Rossi and Spagano, 2018). Finally, scholars have noted that custom gradually gives way to written treaties (international conventions) as a source of international law – another example of substituting customary institutions with broadly conceived designed rules (Lim and Elias, 1997).

All in all, the importance of distinguishing between customary and deliberate mechanisms for developing social rules is evident both philosophically and historically. But how do these two compare? To facilitate a side-by-side comparison, we devise a simple model.

2. Model

This section introduces a formal framework for thinking about social rules. It allows for more structured analysis of the two rule-creation mechanisms outlined above. We start with the following observation: it is now conventional wisdom among law and economics scholars and institutional economists that a mixture of coordination and incentive motives is inherent in most attempts to establish social cooperation based on predictable rules (e.g., Hadfield and Weingast, 2012; Hindriks and Guala, 2015; Bertolini, 2016).

The coordination motive pertains to the fact that the same social dilemma may have more than one cooperative solution. Whether rules can successfully govern the dilemma depends on whether agents expect others to behave according to the same rule. In other words, cooperation requires a logically antecedent common notion of what cooperation means. This common notion can be identified with shared expectations regarding standard responses to specific types of social problems. For example, while the analysis by Coase (1960) suggests that cattle trespass can be addressed through various liability regimes (fencing in or fencing out), any such regime can effectively operate only when ranchers and farmers expect the same liability rule to be applied and enforced by others. The significance of coordination is expounded most prominently in the so-called coordination theory of law, suggesting that the primary function of law is to facilitate coordination among members of society, which ensures that their interactions are orderly and predictable (see, e.g., Postema, 1982; McAdams, 2000; 2009).

In turn, the incentive motive pertains to a different challenge: even when a common notion of cooperation exists, individuals may have insufficient incentives to cooperate. Ranchers who are required to compensate for crops destroyed by trespassing cattle may be unwilling to do so because they expect too little punishment for non-compliance. The provision of incentives to cooperate requires punishments against non-cooperators, e.g., in the form of boycotts of defecting parties. This, in turn, might introduce a higher-order incentive problem: those who are supposed to punish may lack sufficient incentives themselves (see, e.g., Axelrod, 1986).

However, the coordination and incentive problems are not the only challenges with developing social rules. Cooperative equilibria are not merely "out there" waiting to be selected. In more plausible scenarios, the feasibility of various rules of social order is contingent, in part, on the private characteristics of the agents involved, such as their preferences regarding different outcomes or value attached to the future relative to the present moment. When these characteristics are not known at the outset, an additional challenge is identifying the equilibria – i.e., determining which patterns of cooperative group action can be supported by sufficient incentives. For example, among the conceivable methods of regulating liability for cattle trespass, some may not be feasible. Certain liability rules may be overly inconvenient or costly for enforcers to implement, with the cost of their implementation depending on private characteristics of individuals. Therefore, when agents possess qualities that are relevant to their decision-making, and other agents do not have perfect knowledge of these qualities – technically speaking, when information is incomplete – establishing cooperative rules presents a particularly difficult challenge. This issue of finding out whether a rule is or is not an equilibrium is referred to as the discovery challenge.

The model incorporates the incentive problem, the coordination problem, and the discovery challenge. In the model, agents can exert effort that is socially beneficial if effort is exerted collectively. However, only one agent benefits from collective effort at any given time, creating a temptation for others to defect – hence the incentive problem. Moreover, agents face different versions, called *variants*, of the same underlying social dilemma, and each agent is interested in cooperating only in some of them. To achieve any benefits from cooperation, agents need a commonly understood concept of which variants should be collectively addressed: a shared notion of desirable and undesirable behavior. Because many possibilities may exist, this poses a coordination problem. Finally, to account for agent-specific factors that influence decision-making, the model assumes that agents' preferences regarding which variants of the social dilemma they wish to address collectively are private information. Therefore, to cooperate based on any such shared notion,

they need to determine which cooperation possibilities are feasible, i.e., overcome the discovery challenge.

2.1.Basic design

We focus on a stylized scenario where agents repeatedly face a social interaction involving a mix of an incentive problem, coordination problem, and discovery challenge. For simplicity, we assume the group consists of only two agents: Agent 1 and Agent 2, though the logic can be extended to larger groups. The social interaction is structured as follows: agents have a choice of exerting effort or not in every period. One agent, referred to as the affected party, may gain utility benefits if both parties exert effort. The other agent, called the contributing party, does not gain utility benefits in the current period regardless of agents' actions. Effort is costly: whoever makes effort incurs a utility cost c. Moreover, agents' roles change randomly. In one period, an agent may be an affected party and become a contributing party in another. In each period, the roles are reassigned, with each agent having a ½ probability of taking on the affected party role.[5]

The social interaction may be interpreted in many ways. For example, the affected party may be a buyer wronged by a seller, and the wrongdoing can be rectified if two agents pool their resources and risk a costly confrontation with the seller. Alternatively, the affected party may benefit from retaining a resource long-term, provided that both agents refrain from consuming or using it immediately. In both scenarios, it is assumed that a single agent is too weak, incompetent, or otherwise unable to secure the benefit alone.

Importantly, the social interaction is not identical each time it occurs but has multiple possible variants. For instance, the seller's wrongdoing experienced by the affected party may involve delivering a product of defective quality, insufficient weight, or non-standard granularity; exploiting the resource may take several forms, each inhibiting different long-term uses while having minimal impact on others, etc. Formally, let $\Omega = \{1, \dots, \mathrm{N}\}$ denote the set of all conceivable variants of the social interaction. The specific variant that the agents face in any given period is randomly drawn from a uniform distribution over Ω at the start of that period. At the current stage of the argument, the set Ω is fixed. In the discussion section, we

[5] Up to this point, this setup combines key features of assurance games and broadly conceived social dilemmas. On the one hand, lack of effort is a risk-free action. When any agent refuses to exert effort, the cost of contribution is not incurred and the agent's outcome does not depend on the other party's choice. The existence of a risk-free option relates our formal framework to assurance games. On the other hand, the cooperative option is burdened with an incentive problem: because effort does not benefit all agents simultaneously, one party contributes in hope that it will benefit in the future. This party faces a temptation to defect and avoid the cost of effort in the current period. The temptation makes our scenario into a social dilemma.

will explore the role of rule complexity by considering higher or lower granularity within the possible variants.

Variants matter because they affect the payoffs. Agents benefit from mutual effort exerted in some variants of the social interaction but not in others: they may be wronged by some misbehaviors by sellers and indifferent to other misbehaviors, they may value some long-term uses while disregarding others, and so on. Whether an agent benefits from mutual effort in a given variant is a characteristic specific to that agent. We will say that an agent who benefits from mutual effort in some variant of the underlying social interaction *cares* about that variant. The agent-specific preference scheme that determines whether Agent i cares about a specific $k \in \Omega$ will be called this agent's classification scheme. Formally, a classification scheme is the set of all variants about which Agent i cares and will be denoted as Ω_i.

		Contributing party: Effort	Contributing party: No effort
Affected party: Effort		$u_k - c$ / $-c$	$-c$ / 0
Affected party: No effort		0 / $-c$	0 / 0

Figure 1: Stage-game payoff matrix – variant k

Figure 1 shows the payoff matrix for variant k. As is conventional, the top line indicates payoffs of the row player, while the bottom line shows those of the column player. The variable u_k represents the affected party's valuation of mutual effort exerted in variant k of the social interaction. u_k can take the value g when Agent i cares about k, or 0 otherwise. We assume that $g > 2c$, so that mutual effort in cases in which agents care about k is socially beneficial, meaning that exerting effort is a cooperative behavior.

Classification schemes are not only specific to agents but also private. Agents do not know each other's classification schemes with certainty. Instead, they hold beliefs. $\beta_i^k(t)$ will denote the belief of Agent $i = 1, 2$, held at the beginning of period t, that Agent $j \neq i$ cares about variant k. As previously mentioned, this assumption of incomplete information corresponds to the familiar and universal feature of the real world that individuals have only limited knowledge of others'

private values and objectives. However, this knowledge is dynamic; when agents observe others' actions, their beliefs are accordingly updated.

Finally, we make the simple assumption that $\delta \in (0,1)$ is an inter-period discount factor common to both agents. The events of each period unfold as follows:

1) Some variant k of the social interaction is drawn from a uniform distribution over Ω and one of the agents, each with probability ½, becomes the affected party;
2) The game represented in Figure 1 is played;
3) Payoffs are distributed and the period ends.

In any period T, an agent maximizes the expected utility over the remainder of the game, represented by $\sum_{\infty}^{t=0} \delta^{T+t} v_{T+t}$, where v_{T+t} is the expected utility in period $T+t$.

Regardless of which specific interpretation of the model is chosen, it is based on the necessity of collective action as its conceptual foundation. It assumes the absence of any distinct external enforcement agency, such as state enforcement organizations, and the impossibility of externally manipulating agents' payoffs. Cooperative outcomes can only be achieved through the direct involvement of the parties in the social interaction.

The reliance on collective action is a significant assumption, but it can be justified both historically and from the abstract legal theory perspective. Historically, collective action within security networks – such as the cognatic groups, lineages, clans, or voluntary alliances – served as the primary means of enforcing rights, freedoms, and other legal positions in pre-modern societies (see, e.g., Unger, 1976; Murray, 1977; Westbrook, 1988). The extensive involvement of specialized and organized enforcement agencies in law enforcement is a relatively recent phenomenon, dating back only to the beginning of the industrial era (Friedman, 1995). Additionally, owing to the anarchic structure of the international system, collective action has historically been and still remains the key enforcement mechanism in international law (Hathaway and Shapiro, 2011). This suggests that analyses assuming collective action as the sole available means of enforcement are applicable to a wide range of historical and contemporary issues.

More abstractly, it can be argued that all enforcement, even that which is supposedly centralized and organized, requires some degree of coordination. For instance, Basu (2018) suggests that because enforcers are always numerically many, enforcement action must inherently involve collective action. Thus, the difference between state-enforced rules and socially enforced rules is a difference of degree rather than kind. Similarly, Postema (1982) contends that the coordination provided

by law operates on three levels: coordination between regular agents, between regular agents and representatives of the legal machinery, and among the representatives of this machinery, including enforcers.

2.2.Complete information and payoff-dominant equilibria

Irrespective of the state of agents' knowledge, the repeated game has one obvious equilibrium: no cooperation. A strategy of no cooperation simply means always opting for no effort, regardless of the other party's actions. This strategy guarantees an expected per-period payoff of at least 0. Importantly, the no-cooperation strategy is risk-free; it avoids certain utility losses in the name of obtaining uncertain utility gains, now or in the future. It is also evident that no cooperation is a best response to itself: if one agent never contributes, the other agent has no incentive to contribute either. Therefore, a pair of no-cooperation strategies constitutes an equilibrium, irrespective of the agents' beliefs about each other's private classification schemes.

However, cooperative equilibria may also exist, and when they do, they are preferable to the no-cooperation equilibrium in terms of utility (i.e., they payoff-dominate the no-cooperation equilibrium). To illustrate, consider a complete information scenario where the contents of Ω_i for $i = 1, 2$ are public knowledge. There is no discovery challenge in this scenario: agents know each other's private classification schemes and thus are aware of how their actions affect each other's utility. This means that viable patterns of mutual contribution (i.e., possible equilibria) are also known. However, equilibrium play remains ambiguous because multiple equilibria are conceivable, and the coordination problem persists. For example, agents might cooperate in every variant of the social interaction they both care about, and only those. Alternatively, they may trade effort in some variants for effort in others. In order for an equilibrium to materialize, parties need to have a common notion of "right" and "wrong" behavior – this is, a common expectation of when effort is required.

Limited knowledge about agents' private classification schemes – or more broadly, about the game's structure – adds an important element to the picture: the common notion of "right" and "wrong" needs to develop over the course of the game. We will now systematically consider two ways of establishing such a notion: custom and design.

3. Past-oriented rule creation: Customary rules

Custom can be interpreted as a "spontaneous" mechanism for creating rules. Customs consist of unwritten rules and conventions that emerge organically within

a group without external coordination. They develop over time from repeated interactions among individuals. Custom relies on past experience; it is derived from precedents that have been disseminated in society and solidified into consistent patterns, eventually becoming established as commonly expected standards of conduct.

3.1.Reciprocity

Theoretical studies of norm emergence emphasize the expectation of long-term reciprocity as the key factor for creating and sustaining cooperation when no third-party authority is in place (see, Axelrod and Hamilton, 1981; Fon and Parisi, 2003; Elster, 2011). Through reciprocity-based cooperation, agents trade like-for-like activities in such a way that their exchange is stochastically symmetrical in the long run (see, Fon and Parisi, 2003). For example, automatic reciprocity is considered one of the primary meta-rules in international law (Guzman, 2008).

We begin by evaluating how strong the expectation of reciprocity must be to incentivize effort. Assume that the agents are facing variant k of the social interaction in the current period. Consider the following expectation held by Agent i, being the contributing party: with probability p_k, the affected party will contribute in every instance of variant k, provided that the contributing party continues doing the same; with probability $1 - p_k$, the affected party will not contribute. A rational Agent i will exert effort in the current period if they care about variant k and the following condition is satisfied:

$$-(1-\delta)c + \delta p_k \frac{g-2c}{2N} > 0 \tag{1}$$

The left-hand side represents the contributing party's expected utility value of exerting cooperative effort on behalf of the affected party in the current period. It consists of two components: the cost of contributing incurred in the current period and the continuation payoff, which will be realized with probability p_k. The continuation payoff reflects the expected value accumulated over all future periods in which the agents cooperate in variant k of the social interaction, which constitutes a fraction $\frac{1}{N}$ of all periods. In half of those periods, Agent i is the affected party and enjoys a utility gain g. Conversely, the right-hand side of Inequality (1) is nil; it represents a scenario where Agent i refuses to exert effort and thus the agents never cooperate in variant k in the future.

Condition (1) gives the following specification of the minimum p_k, i.e., the minimum belief by Agent i that the other party will infinitely reciprocate effort when the agents face variant k:

$$p_k > \frac{1-\delta}{\delta}\frac{2Nc}{g-2c} \equiv p_k^* \quad (2)$$

3.2.Signaling and extended cooperation

However, condition (2) is relevant only if it simultaneously holds for *both* parties. If it holds for one agent but not another, only the first one will be willing to exert effort; after the effort is not reciprocated, cooperation breaks down. Does it mean that in asymmetrical cases – i.e., when one party believes that there is a potential for cooperation but the other party does not share this belief – cooperation cannot materialize?

In fact, it can, provided that agents can update each other's beliefs by signaling their intentions to cooperate. Signaling is a method of starting mutually beneficial endeavors well known from the law and economics literature. For example, the literature on norm entrepreneurship exploits this idea. According to this theory, an agent may take the risk of starting a new group practice in hope that others would follow suit. In case they do, future benefits accruing to the originator of the new practice would outweigh the cost of risk-taking. Such first movers, or norm entrepreneurs, often exhibit superior situational awareness, technological knowledge, or another kind of comparative advantage that makes them particularly suitable for ushering in new modes of conduct (Ellickson, 2001). Likewise, Leeson (2008) shows that heterogeneous agents can use constly signals to manifest their intention to cooperate in the long run and thus reduce the "social distance" that impedes beneficial exchanges.

We begin with the simplest case of signaling. Assume that Agent i does not believe that Agent j will exert effort in variant k of the social interaction in the current period. However, he also believes that with probability p_k, once Agent j observes Agent's i effort, he will cooperate in all subsequent periods in which variant k occurs.

In other words, Agent i may signal the intention to cooperate in variant k of the social interaction in the future by unilaterally exerting effort in the current period. With probability p_k, this signal will initiate reciprocity-based cooperation in all future instances of variant k. Signaling involves incurring the cost of effort twice: first, in the current period as a costly signal, and then in the next period in which agents face variant k again. The outcome of this next period will reveal whether signaling has proven successful. With probability p_k, both agents will exert effort from this next period onwards. Conversely, with probability $1-p_k$, Agent j does not respond to the signal and exerts no effort in the next period instance of variant k. Therefore, the following criterion must be met for a rational Agent i to engage in signaling:

$$-c + p_k \frac{\delta}{1-\delta} \frac{g-2c}{2N} - (1-p_k) \frac{\delta c}{\delta + (1-\delta)N} > 0 \quad (3)$$

The left-hand side of Inequality (3) represents a signaling scenario. In the current period, Agent i incurs the cost of effort c. This opens two possible futures. With probability p_k, Agent j will cooperate in variant k in all future periods, resulting in a discounted continuation payoff $\frac{\delta}{1-\delta}\frac{g-2c}{2N}$ for Agent i. However, with probability $1-p_k$, Agent j refuses to make effort the next time variant k occurs. In this case, Agent i incurs the cost c one more time and, upon realizing that Agent j does not cooperate, ceases to contribute. The next occurrence of variant k is discounted by the factor $\frac{\delta}{\delta+(1-\delta)N}$. The right-hand side of Inequality (3) represents the scenario where Agent i decides not to signal.

Inequality (3) yields the following condition specifying the minimum p_k that makes signiling rational:

$$p_k > \frac{1-\delta}{\delta} N \left(1 + \frac{4\delta c - (g-2c)(\delta + (1-\delta)N)}{(2-2\delta)Nc + (g-2c)(\delta + (1-\delta)N)}\right) \equiv p_k^{**} \quad (4)$$

Condition (4) is stricter than Condition (2) analyzed previously. It can be shown that as long as the discount factor $\delta < 1$ and $p_k^* \leq 1$ – meaning both parameters remain in their meaningful ranges[6] – the minimum value p_k^{**} exceeds p_k^*. This result is intuitive: initial signaling is a costly way to initiate cooperation. However, unlike reciprocity-based cooperation, it does not require symmetrical expectations. Therefore, Condition (4) only needs to be satisfied by the agent who assumes the role of risk-bearer.

This conclusion squares with the conventional account of norm entrepreneurship that can be encountered in the literature. This conventional account highlights that individuals take risks by initiating changes to prevalent social practices with limited initial expectation of success. For this reason, successful norm entrepreneurs tend to exhibit comparative advantages in taking such risks, e.g., they have greater stakes in social change, have the capacity to accommodate the associated costs, or possess information unavailable to others. We represent this comparative advantage as an advantage in knowledge: an unusually strong belief p_k that others will eventually adopt new, cooperative modes of behavior if they observe first originators.

[6] $p_k^* > 1$ means there are insufficient incentives to sustain cooperation even when agents are fully confident that effort will be reciprocated.

The logic of signaling can be extended further. While our framework outlines a specific type of signaling procedure, the concept can be generalized to encompass all costly signaling methods aimed at identifying long-term cooperation opportunities. For instance, agents may trade efforts in variants of the social interaction that only one agent cares about for efforts in variants that the other agent cares about. Naturally, the hypothetical participation thresholds $p_i^{(a)}$ for singnaling that would convey the necessary information would be even higher than p_i^{**}. In other words, the idea is straightforward: singnaling procedures become more costly with increased complexity of the underlying social interaction, as measured by the total number of possible variants n.

3.3.Equilibrium

An equilibrium in which agents organically create customary rules can be imagined as a superposition of reciprocity-based cooperation and cooperation initiated through signaling. How the rule-creation process unfolds depends on the structure of initial beliefs (i.e., $\beta_i^k(1)$ for $i = 1, 2$ and $k = 1, \dots, N$); in other words, the extent of cooperation achieved through the formation of customary rules depends crucially on the initial *familiarity* with each others' private preferences and objectives. This degree of this familiarity determines whether, when variant k occurs for the first time, agents will engage in reciprocity-based cooperation, signaling, or disregard the possibility to contribute altogether. Those initial responses shape the scope of customary rules and define the content of the common notion of "right" and "wrong" that guides collective action.

When $\beta_1^k(1) > p_1^*$ and $\beta_2^k(1) > p_2^*$, reciprocity-based cooperation may begin. The agents are sufficiently convinced that they can both benefit from exerting mutual effort in variant k. Once cooperation begins, agents update their beliefs $\beta_i^k(t)$ to 1, allowing cooperation to continue.

In turn, in asymmetrical cases when $\beta_j^k(1) \leq p_j^*$ but $\beta_i^k(1) > p_i^{**}$ for $i \neq j$, Agent i faces an incentive to signal. For those combinations of initial beliefs, one agent unilaterally exerts effort when variant k occurs and thus signals an intention to cooperate in the future. This signal causes the other agent to update their belief $\beta_i^k(t)$ to 1. After the belief is updated, Condition (2) is satisfied, allowing reciprocity-based cooperation.

If any parties finds additional ways to signal cooperation opportunities, these ways are also used if similar conditions in the form $\beta_i^k(1) > p_i^{(a)}$ are met. In general, reciprocity has the lowest threshold of initial familiarity with other parties' private peferences and objectives; signaling that enables reciprocity has a higher threshold; and more complex signaling has even higher thresholds. Finally, when

the beliefs are insufficient for agents to rationally engage in any form of signaling, customary rules do not develop.[7]

For the sake of simplicity, if we restrict the equilibrium to include only reciprocity and basic signaling, the equilibrium strategies and beliefs can be characterized as follows:

Agent's i $(= 1, 2)$ strategy:

1) If variant k occurred for the first time, $k \in \Omega_i$, and Agent i is expected to exert effort, Agent i makes effort; otherwise, they ignore variant k;
2) If variant k occurred at least once in the past, $k \in \Omega_i$, Agent i is expected to exert effort, and all agents who were expected to make effort in the past when k was occurring did so, Agent i exerts effort; otherwise, they ignore variant k.

Agent's i $(= 1, 2)$ beliefs $\beta_i^k(t)$ for $\omega_k \in \Omega$:

1) $\beta_i^k(t) = 1$ if Agent $j \neq i$ always succeeded in exerting effort in variant k when expected to do so until period $t - 1$;
2) $\beta_i^k(t) = 0$ if Agent $j \neq i$ failed to exert effort in variant k at least once when expected to do so until period $t - 1$;
3) $\beta_i^k(t) = \beta_i^k(1)$ otherwise.

Although the equilibrium is technical, it can be interpreted in broader and relatively simple terms. The lifecycle of customary rules can be divided into two phases. Initially, before each possible variant in Ω occurs more than twice, a trial-and-error process unfolds, which can be referred to as the formative phase of customary rules. During this phase, agents engage in trial-and-error procedures to establish mutually beneficial long-run patterns of group behavior through reciprocity or signaling. They search for cooperation possibilities by revealing private information about their preferences to each other. However, there is a significant limitation on the amount of trial-and-error activity that agents are willing to engage in during this formative phase. The extent of trial-and-error depends on the overall complexity of the social interaction the agents face. In more complex interactions, where specific types of situations (i.e., individual variants) are rarely repeatedare rarely recurring, it becomes more challenging for a robust common

[7] Importantly, agents also update beliefs when someone who is expected to contribute in any given period *fails* to do so. For example, when $\beta_1^k(1) > p_1^*$ and $\beta_2^k(1) > p_2^*$, both agents are expected to exert effort when they face variant k. A failure to contribute when expected conveys private information about the failing agent – i.e., that this agent does not care about variant k.

notion of "right" and "wrong" to emerge spontaneously. This difficulty is reflected in the increasing values of $p_i^{(a)}$ as n increases.

The second phase of the lifecycle of customary rules can be termed the long-run equilibrium. In this phase, no additional private information is revealed, and beliefs remain no longer change. Agents systematically cooperate in those variants of the social interaction in which successful cooperation has been established in the past; the common notion of "right" and "wrong", or a public standard of acceptable and unacceptable behavior, is thus fully developed and stabilized. The remaining variants receive no collective responses and are infinitely ignored. Importantly, the equilibrium involves punishment of opportunistic defectors. Agents who were expected to exert effort but failed to do so are penalized: in a "grim trigger" manner, non-contributors in variant k would be infinitely denied assistance when they become the affected party in the same variant. This incentive prevents contributing parties from shrinking.

This equilibrium idea can be generalized. Technically, it constitutes a special case of a well-known equilibrium design for infinite games with incomplete information. In this type of equilibrium, players reveal private information to others only within a finite number of periods. After a finite number of periods, the game continues stably without beliefs being updated any further (see, e.g., Koren, 1992; Pęski, 2014). The formative phase in the current framework is analogous to the first part of the above-mentioned equilibrium. As suggested, it can be abstractly envisioned as a phase in which agents attempt to discover mutually advantageous cooperation possibilities that would be exploited in the long run. The discovery is made through trial-and-error procedures where agents use costly signals, thus communicating attractive cooperation avenues. Eventually, these endeavors result in the creation of a commonly understood definition of "right" and "wrong", i.e., a public standard of behavior that agents are expected to follow.

4. Deliberate rule design

Design is a method of establishing rules distinct from custom, primarily due to its deliberate nature and the involvement of a third party. According to the coordination account of law, legal rules function as mechanisms of third-party coordination in agents' strategic interactions. They represent an institutional resolution in which "the solution to the coordination problem is sought not by the parties (…) but by an uninvolved third party" (Postema, 1982:184). This third party's role is to preordain "a solution by going through the process of reasoning which the parties should have used in the situation" (Postema, 1982:183-4). In other words, legal systems replace private classification schemes with a unique, third-party public classification

scheme provided by the legal system itself and recognized as salient by social actors.[8]

Legislation is a prime example of third-party coordination. It provides a unified ex-ante characterization of the decision-making logic that should govern a specific class of social interactions. For example, a legislated statute may state liability rules for various types of animal trespass. With these statements in mind, ranchers and farmers can develop shared expectations regarding the appropriate actions to take in future instances of animal-inflicted crop damage.

Our analytical framework represents designed rules as a public classification scheme preannounced before the game unfolds. The act of preannouncement is "public" in the sense of being public knowledge: the content of the classification scheme is known to all agents, all agents know that other agents know it, etc. In other respects, this classification scheme is no different from the private schemes that correspond to private preferences of individuals. Formally, the public classification logic scheme be represented as a subset Ω_L of all variants, which characterizes those variants in which the scheme demands effort.

4.1.Incentives to use designed rules

Intuitively, an equilibrium with third-party coordination requires that the agents believe that the public classification scheme is sufficiently aligned with their private preferences. This means that it demands effort in sufficiently many variants about which agents care and in sufficiently few about which agents do not care. Consider the situation of a farmer who prefers strict liability rules for cattle trespass. Whenever cattle enter a farmland, a cattleowner should be liable to compensate for the damage inflicted. However, the applicable statute consists almost exclusively of negligence rules, and the stipulated negligence threshold is set high. In other words, the rule designer determines that a cattle owner needs to take modest precautions against cattle trespassing (e.g., by erecting a frugal and fragile fence) in order to be relieved from liability. In many cases where cattle damages the crops while fences are in place, no compensation for the farmer is called for. In this example, the classification scheme embodied in the legislation is poorly aligned with farmer's private preferences.

The alignment between a public classification scheme and agents' private preferences can be measured with two magnitudes: the scope of the public classification scheme and its convergence with agents' private preferences. The

[8] Ultimately, the coordination account of law considers legal systems elaborate conventions. Those conventions assign a special status to classification schemes originating from sources socially regarded as "legitimate" while disregarding other classification schemes as contingent and thus irrelevant.

scope is the fraction of all possible variants in which the public classification scheme demands effort; we write s_L to represent this share. Convergence, on the other hand, indicates how often the scheme demands effort in variants that agents care about. This can be captured by a single parameter, v_i, $i = 1, 2$ representing the fraction of variants in Ω_L that Agent i cares about (and for which effort is requested by the public classification scheme because they are in Ω_L). Thus, the share $s_L v_i$ represents the portion of *all* variants that are deemed relevant both by Agent i and the public classification scheme.

It is now possible to characterize the incentives to use third-party coordination, i.e., to contribute whenever the public scheme demands. We call this behavior Ω_L-directed cooperation. The incentives to participate in Ω_L-directed cooperation will be characterized in several steps. First, consider the perspective of Agent i who assumes that Agent j would always partake in Ω_L-directed cooperation, provided that Agent i never fails to do so. Imagine the following scenario: variant k occurs, Agent j is the affected party, and the public classification scheme Ω_L demands collective effort in variant k of the social interaction. In this case, a rational Agent i has an incentive to participate in Ω_L-directed cooperation in the current period if:

$$-(1-\delta)c + \delta s_L \left(\frac{v_i g}{2} - c\right) > 0 \tag{5}$$

Expression (5) has a straightforward interpretation. The right-hand side represents a situation where Agent i does not exert effort in the current period. Consequently, Agent j ceases to participate in Ω_L-directed cooperation and third-party coordination collapses.

The left-hand side of Expression (5) represents the expected utility from participating in Ω_L-directed cooperation. This can be divided into two parts: the cost of participating in collective action during the current period, and the continuation payoff, which is the expected utility achieved throughout all future periods. In the current period, Agent i is required to bear the cost of effort c. Subsequently, Ω_L-directed cooperation continues infinitely, with Agent i expecting to receive a per-period utility of $s_L(\frac{v_i g}{2} - c)$. This is because Agent i knows that the public classification scheme identifies the share s_L of all variants as requiring collective action. Every time one of these variants occurs, agents collectively contribute, which entails a cost c to Agent i. Moreover, Agent i cares about a fraction v_i of these variants, and since Agent i assumes the role of the affected party on average in every second period, this yields an average utility gain $\frac{v_i g}{2}$.

Note that $s_L(\frac{v_i g}{2} - c)$ can be negative. In such a case, Condition (5) will never be satisfied, and thus Agent i will never participate in Ω_L-directed cooperation. If $s_L(\frac{v_i g}{2} - c)$ is positive, Condition (5) translates into the following condition specyfing the minimum convergence between the public classification scheme and Agent's i private preferences:

$$v_i > \frac{2c}{g}\frac{1-\delta(1-s_L)}{\delta s_L} \equiv v_{min} \tag{5.1}$$

Expression (5.1) specifies a participation threshold. The value v_{min} is the minimum convergence between the public classification scheme and Agent's i private preferences that justifies Agent's i participation in Ω_L-directed cooperation, provided that the other agent participates as well. The participation threshold has intuitive properties: agents with higher δ, meaning higher value attached to future utility relative to present utility, have lower partipation threshold. Likewise, higher benefit g as well as lower cost of effort c also decrease v_{min}.

It is important to reiterate that the participation threshold described in Expression (5.1) is valid only under the strong reciprocity constraint. This threshold specifies Agent's i incentives to participate in Ω_L-directed cooperation, assuming that this agent knows that the other party *already* faces similar incentives. However, it is unrealistic to assume such knowledge in all circumstances. Agents do not have direct access to each other's private classification schemes; this information remains private. As a result, neither agent can directly observe whether the public classification scheme Ω_L aligns sufficiently with *the other agent's* private preferences.

While unable to directly observe each other's private preferences, both agents entertain beliefs about them. We have previously characterized those beliefs by specifying that $\beta_i^k(t)$, $k = 1, \dots, N$ denotes Agent's i belief that Agent j cares about variant k of the social interaction. Agents can also use $\beta_i^k(t)$ to infer the probability of convergence between the public classification logic and private preferences of the other agent. We write $\xi_i(\kappa)$ to denote the function (technically: a cumulative distribution function) that characterizes Agent's i belief that Agent's j convergence level amounts to at least κ. Put differently, $\xi_i(\kappa)$ is a probability, as assessed by Agent i, that the convergence v_j between Agent's j private classification logic and the public classification logic is κ or less.

Beliefs about each other's convergence levels are critical for the possibility of equilibrium with Ω_L-directed cooperation in which both agents participate. They overcome the difficulty associated with the lack of direct insight into the private classification scheme of the other agent. Instead of assuming that the other party

will participate in Ω_L-directed cooperation, agents attach a subjective probability to such a scenario. From the perspective of Agent i, this probability is given by:

$$\pi_i = 1 - \xi_i(v_{min}) \tag{6}$$

Expression (6) represents Agent's i belief that the other agent's private preferences are sufficiently aligned with the public classification logic to engage in Ω_L-guided cooperation. In other words, it expresses, from Agent's i perspective, the probability that Agent j has sufficient incentives to be coordinated by the public classification scheme Ω_L in all future periods, provided that Agent i continues doing the same. Expression (6) is important because it can be used to make Condition (5.1) operational in a world where agents' knowledge of others' private preferences is incomplete. It allows expressing a participation threshold adjusted for imperfections in agents' mutual knowledge.

Assume that the situation is identical to the one characterized when Condition (5) was derived: variant k occurs in the current period, Agent j is the affected party, and the public classification logic scheme calls for exerting effort in variant k. It is rational for Agent i to follow the direction of the public classification scheme Ω_L in the current period, if the following condition is met:

$$-(1-\delta)c + \pi_i \delta s_L \left(\frac{v_i g}{2} - c\right) > 0 \tag{7}$$

Condition (7) is stricter than Condition (5) yet it resembles the latter in all but one respect. Agent i still needs to bear the cost of effort in the current period. However, they now believe that the payoff from the continuation game, i.e., the payoff from Ω_L-directed cooperation in all future periods, is expected to be achieved with probability π_i. With probability $1 - \pi_i$, Agent's j private classification logic is not sufficiently aligned with the public classification scheme, and thus Agent j will be unwilling to follow the directions given by the public classification logic now and in the future.

Condition (7) can be rewritten to obtain the minimum belief π_i:

$$\pi_i > \frac{1-\delta}{\delta} \frac{2c}{s_L(v_i g - 2c)} \equiv \pi_i^* \tag{7.1}$$

4.2.Equilibrium

With Condition (7.1), we can characterize the equilibrium in which both agents adopt the public classification scheme Ω_L, i.e., the set of rules that has been predefined before the game unfolds, to coordinate their behavior in every period.

Agent's i $(= 1, 2)$ strategy:

1) Adapt the actions demanded by the public classification scheme Ω_L (i.e., engage in Ω_L-directed cooperation) if $\pi_i(t) > \pi_i^*$ and if both agents never failed to adapt any of those actions in the past;
2) Always choose no effort if either $\pi_i(t) > \pi_i^*$ is not satisfied or if any agent failed to adapt an action demanded by the public classification scheme Ω_L in the past.

Agent's $i\ (= 1, 2)$ beliefs $\pi_i(t)$:

1) $\pi_i(t) = \pi_i(1) = 1 - \xi_i(v_{min})$ as long as no variant occurred in which the public classification scheme Ω_L demands effort;
2) $\pi_i(t) = 1$ if Agent j always engaged in Ω_L-directed cooperation in the past;
3) $\pi_i(t) = 0$ otherwise.

The equilibrium idea is simple. The third-party classification scheme serves as the common notion of "right" and "wrong." It includes behaviors that should be met with a costly collective reaction. The advantages of Ω_L, as opposed to customary rules, lie precisely in its unique and public character. All agents are presented with the same classification scheme, and all know that this (and no other) scheme has been presented to others and is recognized by each as the valid one. Thus, rule design can resolve the coordination problem: it can overcome the ambiguity associated with selecting one of many possible equilibria.

The question remains whether agents are incentivized to use the third-party classification scheme to coordinate their interdependent actions. Each agent will do so only if they find doing so in their long-term interest, meaning that Condition (5) needs to be satisfied. This interest can be understood as the degree to which the public classification scheme reflects this agent's private preference scheme. However, coordination based on the public classification scheme is possible only if agents believe that both would participate in Ω_L-directed cooperation. Using the third-party classification scheme as a coordination tool depends on the ongoing participation of the other party, which is initially uncertain. Therefore, agents need to take into account the other agent's incentives to participate in Ω_L-directed cooperation. At the outset of the game, they are unsure whether such incentives are in place. Based on their beliefs about the private preferences of the other party, agents are able to derive a belief about the probability that their participation in Ω_L-directed cooperation would be reciprocated.

Third-party classification schemes have an important property: they allow for relatively quick verification of whether agents have sufficient incentives to use them. If agents face incentives to coordinate their actions based on the preannounced concept of "right" and "wrong" embodied in Ω_L, and if they believe that others face similar incentives, they will systematically consult Ω_L when

deciding how to act in response to individual variants of the social interaction. Once Ω_L-directed cooperation begins, each agent becomes confident that others share their interest in sustaining it in the long run. In other words, they have discovered an equilibrium and understand what form this equilibrium will take long-term.

It should also be emphasized that, in addition to the coordination problem and the discovery challenge, cooperation based on designed rules addresses the incentive problem. Agents face an opportunistic temptation: defect in the current period and save the cost of effort, especially if they are in the contributing party role. However, a strong incentive prevents this scenario from materializing. Upon observing a defection by the other party, an agent would infer that their counterpart lacks incentives to engage in Ω_L-directed cooperation. Consequently, the agent would update their belief $\pi_i(t)$ to zero. This would collapse their own incentive to participate in Ω_L-directed cooperation, leading to a breakdown in coordination. To prevent this dismal outcome, rational agents, for whom the participation constraint is satisfied, are not tempted to defect, even when they have nothing to gain from exerting effort on behalf of others. The fear of the decay of the cooperation pattern is a "grim trigger" that effectively prevents defection, thus ensuring the ongoing reliance on predefined rules.

Finally, the same logic suggests that third-party classification schemes are holistic in nature – agents cannot selectively ignore a part of the determinations included in Ω_L without undermining others' confidence in the entire coordination mechanism. This stems from the very structure of the incentives that support the use of third-party classification schemes. The belief that *each and every* variant included in Ω_L will be met with coordinated effort in the future, provided that an agent continues to participate in Ω_L-directed cooperation, is indispensable for the existence of an equilibrium with third-party coordination. Without such a belief, Condition (7) does not hold, and adhering to the classification logic embedded in Ω_L cannot be individually rational. Instead, agents would need to independently discover the cooperative ventures others are interested in pursuing, which would necessitate a trial-and-error discovery process that third-party coordination was designed to avoid.

5. Discussion

In the previous section, the model illustrated custom and design as two distinct rule-making mechanisms. Customary rules arise and evolve spontaneously, with agents' understanding of desirable and undesirable behavior emerging from the past actions of the involved parties. Operating with limited mutual knowledge, parties engage in trial-and-error processes to define a common notion of "right" and

"wrong." This shared notion eventually leads to coordinated effort in situations where collective action has previously been identified as beneficial. In contrast, designed rules provide a common understanding of desirable and undesirable behavior upfront to all parties. A unique, predefined decision-making logic serves as an external coordination mechanism. The systematic use of this logic can also create a cooperative equilibrium: agents follow it to coordinate their efforts because they fear a collapse of coordination in the long run.

Customary rules, social norms, and similar rule creation mechanisms in have been previously modeled in the law and economics literature. However, our approach distinguishes itself from earlier contributions in several respects. Unlike most earlier models, our framework assumes incomplete information, representing an environment in which agents are uncertain about the private preferences and objectives of others. This situates our work within the literature on norms that emerge under conditions of "private assessment" (see, e.g., Okada et al., 2018), a literature that until recently has been almost exclusively anchored in theoretical biology.

Trial-and-error representations of norm emergence have been developed within the framework of evolutionary game theory (e.g., Sugden, 1986; Young, 2001; 2015; Aoki, 2001). These frameworks often depict norms as "sticky" behavioral predispositions that remain constant even when new situations arise or fresh information becomes available.[9] In these models, norms, much like genes in biological populations, compete with each other, being transmitted across generations at varying rates. In contrast, our model adopts a rational actor perspective. Unlike the fixed behavioral schemas assumed in evolutionary models, our approach allows for agents who dynamically adjust their actions based on what they consider optimal at any given time.

The emphasis on agents' limited and uncertain mutual knowledge makes it possible to represent the idea that customary rules are not just any equilibria of repeated games; rather, they need to originate as practices initiated by specific agents at some point in time. While this moment can be difficult to identify in historical records, the formative episode of customary rules is indispensable and cannot be ignored. It plays a critical role in shaping long-standing group practices, that is, the long-term equilibrium.

Likewise, the depiction of designed rules as a tool for third-party coordination builds upon earlier contributions within the coordination accounts of law. Our analytical framework is inspired by the semi-rigorous formalizations by

[9] A similar notion of social norms as relatively constant behavioral predisposition can be found in empirical institutional economics. For a survey, see, Voigt, 2024a; 2024b.

Postema (1982) and McAdams (2000, 2008). To an even greater extent, it follows the repeated game model used by Hadfield and Weingast (2012). However, the approach used in this paper differs in several respects. First, we envision social rules as patterns of behavior where actions deemed socially desirable or undesirable are systematically met with collective responses, such as mutual assistance or collective punishment. In contrast, the model by Hadfield and Weingast conceptualizes rule-based social order primarily as a scenario characterized by the threat of punishment, which functions off-equilibrium. In their framework, the emphasis is on credible deterrence, where the mere possibility of punishment for rule violation is sufficient to maintain social order.

In essence, we find that both customary and deliberate mechanisms for rule creation produce a social order based on rules. Within this order, agents' actions systematically follow a designated pattern, with both agents collectively responding to certain types of events while ignoring other types. Moreover, their actions are mutually understood to be based on a shared normative notion. In the case of customary rules, this notion evolves historically; in the case of designed rules, it is a unique normative classification provided upfront. Nonetheless, the two mechanisms are qualitatively different: each offers unique advantages and has specific limitations that will be examined in this section. The comparison focuses on four aspects: the ability to deal with complexity, path dependency, susceptibility to manipulation of the rule content, and susceptibility to error.

5.1.Complexity

There are multiple ways to understand the complexity of social rules. One way to interpret complexity is as internal differentiation: a rule is considered complex if it distinguishes among cases based on a multitude of independent factors. Conversely, simpler rules group cases into fewer, larger bundles (see Schuck, 1992). For example, tort law: strict liability is a relatively simple liability rule because it assigns liability based on a single criterion of causality. Negligence is more complex as it additionally considers care, treating cases where care has been exercised differently from those where it has not. Finally, comparative negligence is even more complex, differentiating between cases based on two levels of care, each exercised by one of two parties. Within the formal framework used in the previous two sections, this kind of rule complexity can be identified with the number of possible variants N of the social interaction. This measure reflects the degree to which individual cases are bundled: greater bundling (meaning a lower number of potential variants) corresponds to reduced complexity.

Thus, the initial conclusion from the analysis in the preceding sections is that customary and designed rules differ in their capacity to handle complexity. Our argument is straightforward: the creation of customary rules tends to promote

simplicity, while designed rules are capable of managing greater complexity, limited primarily by the cognitive capacities of the involved parties. The view that customary rules are generally simple finds support in existing scholarship (e.g., Alexander, 2007; Bicchieri et al., 2023) and is consistent with our theoretical exposition. As shown in the preceding sections, in the long-term equilibrium where agents develop customary rules, the number of variants of the social interaction that will be collectively addressed depends inversely on the total number of variants, N, as specified in Conditions (2) and (4).

The rationale can be easily understood: a higher N means that individual instances of specific variants k occur less frequently. This reduces continuation payoffs, i.e., the aggregate payoffs from cooperating in future occurrences of variants k. Since the initial investment required from the parties to establish long-term cooperation in any individual variant k – whether in the from of risking reciprocity for the first time or the more complicated from of signaling – remains constant, increased rule complexity makes this investment less rewarding. In other words, simple rules have a lower expected cost of being established and spread in society: they are cheaper to form, communicate, and learn.

Therefore, when faced with challenges posed by complex realities, agents may be more successful in developing customary rules governing their interactions by reducing rule complexity. As Smith (2009) noted, community customs often evolve in informationally efficient ways by favoring repeatable practices that require minimal situation-specific knowledge. This evolutionary tendency is exemplified by the emergence of the social norm of strict liability for cattle trespass among ranchers and farmers in Shasta County, California (Ellickson, 1991). Implementing more complex negligence rules would require information about situation-specific circumstances, such as the precautions taken by the rancher or farmer, which would impose significant informational costs on those seeking to learn and apply the same social norm in the future. In contrast, strict liability rules are easier to formulate, communicate, and observe in practice. Thus, Smith argues that custom tends towards simplification:

> "[T]he message to keep off, the default regime of possession, is easier to communicate to more far-flung parties. (…) The general, formal default — here the norms of trespass and exclusion — have a gravitational pull, beyond the benefits in terms of the relative importance of farming compared to ranching. Even where a more nuanced flip of the rule from fencing in to fencing out would be efficient in the narrow sense of maximizing the value of the two activities, in close cases there is an additional

> reason to stick with the general rule of fencing in, because it comes along with the general exclusion regime for ownership of land, which is undemanding from an informational standpoint." (Smith, 2009:27)

However, the same argument also suggests that simplification can lead to the establishment of less efficient rules. For example, a well-known result from the economic analysis of accidents indicates that negligence rules are often preferred over strict liability rules on efficiency grounds (Dari-Mattiacci and Parisi, 2006). In this scenario, relying on simpler strict liability rules may forfeit the efficiency benefits that could be achieved through more fine-grained rules.

Compared to customary rules, deliberately designed rules appear better equipped to accommodate complexity. In principle, they can differentiate between cases across any number of types or classes. As long as the public decision-making logic is announced in advance, allowing the parties to reconstruct the implied common notion of "right" and "wrong," complexity does not emerge as a limiting factor. This idea is corroborated in our analytical framework. The incentives to use designed rules (i.e., Conditions (5) to (7.1)) do not depend on the complexity of these rules but rather on how closely they align with agents' private classification schemes.

The ability of designed rules, such as those found in legislation, to accommodate high degrees of complexity has been used to justify making use of this capacity in crafting legal rules. For example, De Geest (2013) formulates a "N instruments for N problems" principle in contract law by claiming that:

> "[I]f we want to solve – say – 8 different problems (creating incentives for optimal breach, reliance, precaution, mitigation, incentives to reveal unusually high potential losses, incentives to promise carefully, incentives not to opportunistically renegotiate, and optimally allocating risks) we need 8 separate rules or doctrines that each address one of these problems rather than trying to solve them all with the choice of a single remedy." (De Geest (2013:43)

Applying this principle would often require highly case-specific rules that treat individual cases differently based on a large number of efficiency demands. In the example of 8 problems, the law should recognize that cases differ across at least

8 dimensions, meaning at least 8 independent rationales for differentiating rules applicable to individual cases. In other words, a rational legal design should recognize "the strong disadvantages of using fewer instruments" (p. 43) because "compromise instruments (i.e., single instruments that are meant to solve multiple problems) cannot be fully effective at solving two or more problems when there is tension between them." (p. 45)

A similar principle is occasionally observed in legal history. Legislative intervention often leads to increased complexity within social rules that were originally based on custom. A notable example is "blood money," a liability rule for homicide, common in kin-based legal systems in many parts of early-to-mid medieval Europe. The institution of blood money required the offender to pay a specified amount of money – the value of blood spilt – to the victim's kin and allies, effectively ending a feud. Importantly, the compensation amount was typically based on one variable only: the victim's status, giving no consideration to factors like the differing needs of the victim's kin, the perpetrator's ability to pay, or, arguably, even culpability. However, the history of culpability for homicide among Anglo-Saxons (Robinson, 1980) and Scottish feuding practices (Wormald, 1982) suggest that these additional factors began to play a significant role after royal legislative interventions. In the Scottish case,

> "[r]oyal writs of the thirteenth and fourteenth centuries (…) offered protection to the man who had killed in self-defence. (…) And a provision in *Regiam maiestatem* carefully distinguished between the amount of compensation owed when a horseman in a village rode down and killed a pedestrian, and when he killed one by backing his horse into him; compensation for the first (…) was far greater than for the second, when the pedestrian should have been more careful." (Wormald, 1982:111-112)

However, the discussion above needs several qualifications. First, the assertion that rule efficiency can be enhanced by increasing complexity does not imply that greater complexity *always* results in improved efficiency. The relationship between rule complexity and efficiency has been a subject of debate within the law and economics scholarship. Scholars have often concluded that there exists a nuanced balance or "sweet spot" of optimal complexity, beyond which complexity becomes either excessively high or inefficiently low (e.g., Kaplow, 1995; Wright, 2000).

Moreover, our discussion deliberately omits the cost of designing rules. Researchers point out that more case-specific rules are more costly to make because lawmakers face an increasing "difficulty of specifying the contingencies of a complex environment" (Fon and Parisi, 2007:152). Thus, taking the rulemaking cost into account would allow to paint a fuller picture of the costs and benefits associated with deliberate and customary mechanisms for rule creation. However, because we are primarily interested in the *qualities* of rules that may arise under both regimes, designed rules are treated as endogenous, and the cost of rule provision borne by the rule provider is disregarded.

Finally, we abstract from the learning burden associated with the excessive complexity of rules. Agents who are unfamiliar with the intricate network of rules, legal doctrines, and methods of normative reasoning may find it difficult to acquire the necessary knowledge, or they may do so only at prohibitive costs. In extreme cases, the information burden may undermine the common knowledge assumption – that all parties know the rules, know that others know them, and so forth. Since such common knowledge is essential for the equilibrium in which agents coordinate their actions based on a third-party classification scheme, excessive rule complexity may render the equilibrium impossible.

5.2.Efficiency, inefficiency, and bias

Efficiency is a key theme in analyzing social rules from a law and economics perspective. We have previously established that customary rules develop when, and to the extent, agents are able to discover opportunities for mutual advantage. However, this does not mean that every opportunity to enhance welfare is realized during the formative phase of customary rules, leading to Pareto efficiency. As mentioned earlier, the challenges in discovering these opportunities can prevent this. Instead, we argue a simpler point: agents typically create customary rules that benefit all parties involved. Consequentially, customary rules that benefit one party but harm another are unlikely to emerge unless the disadvantaged party has no role in their development, in which case negative externalities are possible.

Negative externalities have been long recognized as one of the primary sources of inefficiency of custom. Parties who do not participate in norm creation or enforcement can suffer disutility when their interest is not taken into account by norm creators (Coleman, 1990; Posner, 2000). For example, a customary norm developed by landowners regarding waste disposal might impose environmental costs on neighboring communities that were not involved in establishing this norm. Nonetheless, as long as only participating parties are considered, activities of customary rule creators tend to produce Pareto improvements. Agents who originate and enforce customary rules can opt out if new rules or changes to the

existing ones would systematically harm their interests. Therefore, the potential to include a rule component that harms a subgroup of the involved agents seems substantially limited.

The institutional realities seem to corroborate this conclusion. Custom is considered one of the main sources of international law, where sovereign international actors develop practices that gradually become recognized as foundational for commonly binding rules (Shaw, 2017). However, the influential (though sometimes controversial) persistent objector doctrine in international law allows each state to voice an objection to an emerging practice during its formative phase. By doing so, the objecting state can exempt itself from being bound by an undesirable custom (Green, 2016). This corresponds to the possibility, identified in our formal framework, of deliberately withholding cooperation during the formative phase of customary rules. Withholding cooperation in variant k of the social interaction during this variant's first and second historical occurrence conveys a message that the agent does not wish for exerting effort in this variant to be considered socially expected in the future.[10] This mechanism allows agents to directly influence the content of customary rules. By refusing to participate in practices they find objectionable, agents ensure that only tacitly agreed-upon practices become established as obligatory customs.

However, even if spontaneously created rules benefit all parties, the important issue remains how these benefits are distributed – i.e., which party gains more than others. Knight (1992) claims that the distributional effects of spontaneously developed institutions depend on the bargaining positions of the involved agents. The party with a more valuable fallback option, meaning that this party can do relatively better if no cooperation takes place, is in a position to shape such institutions in its favor. Technically, this implies an equilibrium selection mechanism: which rule is observed in reality depends on what happens when agents fail to coordinate on one of the mutually beneficial institutional outcomes.

Norms regulating intergenerational transfer of property can serve as an example of Knight's logic. Typically, all family members must contribute to their joint economic well-being, and failing to respect inheritance norms could lead to a collapse of familial cooperation. However, whose interests are favored by these

[10] According to the persistent objector doctrine, objections must be explicitly voiced; silence of a state is insufficient for this state to be recognized as an objector (Green, 2016). Our formal framework cannot perfectly recreate explicitly voiced opposition. The framework includes only two agents who take part in interactions in every period, without inactive observers. Therefore, withholding cooperation in the formative phase, before the other agent develops a solid expectation of cooperation in a specific variant of the social interaction the future, is the only option to communicate that the customary rule should not evolve into demanding cooperation in such cases.

norms depends on the relative bargaining positions of the family members. Children who have sources of income other than the family estate may secure more favorable succession rights, such as a guaranteed share of inheritance for every child. Their fallback option, which they can rely on if familial cooperation breaks down, allows them to make demands more comfortably. Conversely, children whose material well-being throughout their adult lives strongly depends on the family estate are in a weaker position and may be subject to less favorable norms, such as primogeniture, which favors the eldest son – the child whose effort typically has the highest immediate value to the family head.

The model presented earlier does not currently incorporate a similar reasoning, but this could be done after a slight modification. Consider a similar model with three agents instead of two. In the new setup, cooperative outcomes can be achieved in any given period not only when all agents exert effort, as previously assumed, but when the total sum of efforts exceeds a critical mass of ½. Further, assume the agents have unequal contributions to make: Agent 1 can contribute ½ of the total effort, while the remaining two agents can each contribute only ¼. This disparity in effort may result from differences in economic productivity, political power, or other exogenous factors.

Given this disparity, it is natural to expect it is natural to expect that the common notion of cooperative behavior that develops on a customary basis would lean toward one preferred by Agent 1. This is because Agent's 1 effort can be effectively combined with that of *either* of the other two agents, providing a favorable fallback if one of them chooses not to cooperate in some variants of the social interaction. In contrast, Agents 2 and 3 need Agent 1 to achieve cooperation in any variant they care about and have no fallback option. However, if the underlying economic or political factors change, giving more power to Agent 2 and less to Agent 1, the customary rules should adjust accordingly, now in favor of Agent's 2 preferences. An important conclusion is that a change in exogenous economic conditions is both sufficient and necessary to trigger a change in customary rules: these rules respond to such conditions and, after the formative phase is complete, they do not change unless those conditions shift. Only during the formative phase, when the content of the rules is still dynamically molded, can Pareto-improving changes be expected.

On the other hand, designed rules can be exogeneously changed even if the underlying economic forces remain constant. The change occurs by reshaping Ω_L (e.g., through announcing a new legislative act) and thus recreating the unique focal point around which the social notion of "right" and "wrong" are centered. In this way, the incorporation of rule components that favor one party or group but disadvantage another seem to be compatible with deliberate mechanisms for creating rules. This is caused by the already mentioned holistic nature of third-party

coordination. Agents will rely on third-party coordination it if they have *sufficient* (albeit possibly imperfect) incentives to do so, and have no option of selectively rejecting its particular elements.

This opens the possibility of asymmetrical rule manipulation; the content of Ω_L can be changed to align more closely with the interests of one party or group at the expense of another, even when all the underlying economic forces are stable. The possibility of incorporating such asymmetries into designed rules allows for designing rules that improve efficiency in the Kaldor-Hicks sense: total utility is improved, but at least one party is disadvantaged in the process. Thus, when satisfying the preferences of one party or group at the expense of another one is socially productive, design becomes a superior mechanism for rule production from the efficiency standpoint.

Figure 2 graphically illustrates this reasoning. The horizontal and vertical axes of both panels correspond to agents' utility levels from collective action based on various common notions of "right" and "wrong." The left panel represents notions of social wrong implied in designed rules. Here, 0 is the baseline utility level that must be exceeded to sustain Ω_L-directed cooperation (as indicated by Condition (5)). Only those public classification schemes that ensure positive utility can be rationally used as third-party coordination devices. Arrow a_I represents a change in the content of Ω_L that increases the sum of utilities of both parties. However, the change is asymmetrical: Agent's 1 utility decreases but Agent's 2 utility simultaneously increases.[11] For example, in the animal trespass problem, Ω_L can be changed by introducing a requirement that farmers fence their crops. If fencing crops is a cheaper way to prevent crop damage than fencing animals in by ranchers, this change is socially beneficial. However, it also shifts the burden of preventing crop damage from ranchers to farmers; if farmers do not happen to simultaneously be farmers and vice versa, the roles are not reversible and the new requirement systematically favors farmers at the expense of ranchers.[12]

Importantly, asymmetrical changes in the content of Ω_L do not disrupt the equilibrium: both agents still face sufficiently strong incentives to use the third-party classification scheme Ω_L as a coordination mechanism because each expects to achieve some positive utility level.

---

[11] We conceptualize this change as exogeneous and unanticipated; otherwise, if agents expected rules to change, they would not rationally expect their future interactions to be governed by the same rules as the present ones, leading to a decrease in the discount factor δ. While the analytical framework could also represent this scenario, it is left out of the scope of the current analysis for the sake so simplicity.

[12] Naturally, this example assumes that transaction costs are too high for the parties to renegotiatie their positions ex post.

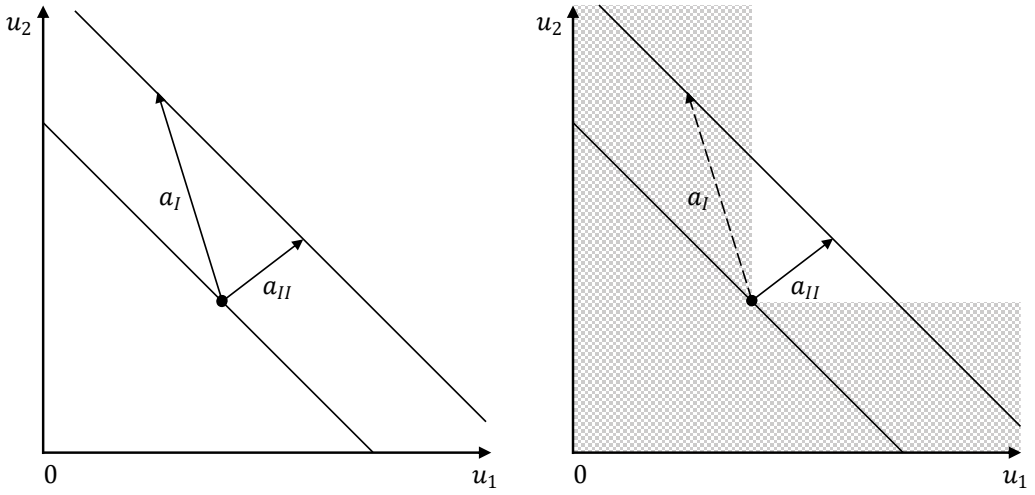

Figure 2. Viable changes in customary rules (left panel) and designed rules (right panel)

On the other hand, the panel on the right-hand side illustrates an analogous change in the content of customary rules. Since agents can reject the change of rules that disadvantage their individual utility, asymmetrical changes do not occur, even if desirable from the Kaldor-Hicks perspective. Therefore, unless the underlying economic conditions change (for example, unless the cost of fencing crops decreases or the value of crops increases), the only viable direction in the evolution of customary rules are Pareto improvements such as the one represented by arrow a_{II}. With this type of change, both agents simultaneously benefit.

Interestingly, a related idea has been voiced by a legal historian. In his description of the first codifications of Greek laws, Smith (1922) acknowledges that drafting of legal codes in Greek colonies (where the first Greek legal codes were made) was frequently an attempt to reconcile conflicting interests of colonists of diverse cultural backgrounds.

> "In many cases codification was imperative because the members of a colony were recruited from different cities. Hence no single set of customary laws could be entirely satisfactory even if it proved suitable to the new conditions. The first codes were made in the western colonies which were farther from Greece than the eastern colonies both in distance and in the difficulty of the voyage." (Smith, 1922:188)

This description squares with our concept of rule design mechanisms that can accommodate a moderate amount of conflicting interests. Conversely, the description suggests that customary rules could not be easily merged to produce a functional social order.

5.3.Susceptibility to manipulation by vested interest

However, the possibility of manipulating the content of designed rules, regardless of the underlying economic fundamentals, also allows for socially inefficient rule manipulation. In particular, it permits manipulation that is inefficient according to the Kaldor-Hicks criterion and even such that makes every agent worse off. This possibility is straightforward: the manipulations represented graphically as arrows a_I and a_{II} in the left panel begin and end at two stable equilibria with third-party coordination. Therefore, the direction of change can be reversed compared to what was previously discussed – i.e., a change from more to less efficient rules is equally possible.

A backward movement along arrow a_I corresponds to a legislative or regulatory capture scenario, where rulemakers devise socially inefficient rules that favor a specific faction or group. In the specific case illustrated in Figure 2, Agent 1 would benefit at the expense of the general social welfare. Thus, assuming that perverse incentives are in place, the holistic nature of designed rules and their relatively loose connection to the underlying economic forces allow a rule designer to serve the vested interests of a subset of agents, even to the detriment of the general welfare.

Beyond a regulatory capture by a subgroup of agents, there is also a hypothetical possibility of rule manipulation that leaves *all* interested parties worse off, as represented by the backward movement along arrow a_{II}. While both agents would suffer a utility loss as a result of such rule manipulation, it would still not undermine an equilibrium with Ω_L-directed cooperation; sufficient incentives to cooperate based on the public classification scheme still exist.

A manipulation of the rule content that leaves all parties worse off may happen when the rule creation mechanism itself becomes captured by the agents operating it – e.g., by legal professionals benefiting from the existence of overly complicated and troublesome rules. The possibility that self-serving representatives of the legal system may, once this system becomes sufficiently bureaucratized and centralized, distort it to the detriment of the general population has been seriously considered by prominent legal scholars. For example, Raz (1994) claims that

> "[A]s a result of the growth of a legal profession and a highly articulated legal culture, legal issues are formulated in technical terms, caught in legal categories which are far removed from the way ordinary people understand their conduct and interactions with others. The law becomes financially inaccessible and conceptually remote and alienating." (Raz, 1994:372)

Likewise, Ogus (2002:434) argues that "[t]o the extent that they have monopolistic power, lawyers can exploit the key features of legal culture to extract rents: the law used can be more formalistic, more complex and more technical than is optimal". Similar arguments are raised by Epstein (2009).

All in all, the very nature of third-party coordination, whose ongoing viability depends on designed rules being *sufficiently* (but not necessarily perfectly) efficient, and which can be manipulated easier than the content of customary rules, invites the possibility of asymmetrical treatment. This possibility is double-edged: on the one hand, it enables Kaldor-Hicks efficiency improvements that are difficult to incorporate into customary rules. On the other hand, it allows socially inefficient manipulation that serves vested interests. In short, the capacity for efficiency that goes beyond the limitations of customary rule-creation mechanisms can simultaneously accommodate significant inefficiency.

5.4.Susceptibility to error

The argument presented in the previous subsection can be also extended to include unintentional errors committed in the design of social rules. A mistake in devising rules may lead to outcomes that differ from, or even contradict, the intended results. While occasionally such design errors may produce socially beneficial outcomes, in most typical cases such "unintended consequences" frustrate the original, socially beneficial intention (see, Baert, 1991).

It can be inferred from the previously presented model that social rules developed through deliberate design are more susceptible to errors resulting in adverse outcomes compared to customary rules. This is because the participatory nature of custom formation allows the interested parties to express objections verbally or communicate them through acts of non-cooperation. This process reveals the dispersed knowledge of potential negative consequences, even if the exact nature of these consequences is not communicated directly. Conversely, deliberate rule-making lacks such feedback mechanisms, which impairs its ability to filter out mistaken ideas, leading to unintended, adverse consequences. In other words, despite the presence of significant errors, third-party coordination based on

externally provided classification logic may still be preferable to completely rejecting this coordination method.

6. Closing remarks

We have presented two highly stylized mechanisms for rule-making, portrayed them in game-theoretical terms, and compared them in light of several criteria. On a concluding methodological note, it must be stressed that the analysis of customary and deliberately designed rules treated them as "ideal types" (Weber, 1949). Ideal types are simplified and purified representations of real-world objects that emphasize their core aspects and disregard circumstantial characteristics. Because of this simplification, it is natural to expect that our stylized model will not perfectly correspond to rule-creation mechanisms observed in the real world. However, this does not mean that they are useless; on the contrary, the analysis of ideal types makes it possible to distill essential elements of real-world phenomena.

We begin from a simple premise that agents with different beliefs need to coordinate on mutual understandings of what constitutes desirable and undesirable conduct, and how this conduct should be responded to. Moreover, we add that these beliefs are often private, which is a consequential fact for the shape and form of cooperation between agents. From these two premises, the paper finds that customary and designed rules tend to differ qualitatively. Customary rules tend to be relatively simple; in general, those cooperative practices that are easy to observe, understand, and replicate have better chances of becoming expected in wider society. Internal differentiation seems to be more compatible with rules that are supplied upfront and with a significant element of deliberate design.

The paper also suggest that different factors drive the efficiency of customary and designed rules. Customary rules are strongly shaped by underlying economic forces, such as the cost-benefit ratio of different activities. When these forces change, spontaneously developed methods of cooperation are expected to adapt accordingly. In contrast, designed rules are less responsive to economic factors. This is because of the ambiguity of incentives to use third-party coordination mechanisms, of which designed rules are a special case. Third-party coordination can function if agents have barely sufficient, though imperfect, incentives to participate. The indivisibility of these incentives presents two options for rule design. First, there is the potential to deliberately optimize rules for greater efficiency in the Kaldor-Hicks sense – something typically absent in customary rules unless the underlying economic forces shift. However, the same indivisibility can also lead to the manipulation of rules in socially inefficient ways, for example

through serving the vested interests of specific social actors or incorporating errors that lead to adverse unintended consequences.

The paper is theoretical; its ambition was to enhance our understanding of social rules from a rational choice perspective. For this reason, its potential applications have only been vaguely indicated in the discussion. Nonetheless, we still believe that our theory can shed light on historical and contemporary contexts where the two stylized institution-making methods function as alternatives. These contexts include, among others, early legal history, international law, and possibly others.

Literature

Alexander, J. 2007. *The Structural Evolution of Morality*. New York: Cambridge University Press.

Aoki, M. 2001. *Toward a Comparative Institutional Analysis*. Cambridge, MA, and London: MIT Press.

Axelrod, R. 1986. "An Evolutionary Approach to Norms," 80 *The American Political Science Review* 1095-1111.

Axelrod, R., and Hamilton, W. 1981. "The Evolution of Cooperation," 211 *Science* 1390-1396.

Baert, P. 1991. "Unintended consequences: a typology and examples," 6 *International Sociology* 201-210.

Barkun, M. 1968. *Law without Sanctions. Order in Primitive Societies and the World Community*. New Haven: Yale University Press.

Basu, K. 2018. *The Republic of Beliefs. A New Approach to Law and Economics*. Princeton and Oxford: Princeton University Press.

Bertolini, D. 2016. "On the Spontaneous Emergence of Private Law," 29 *Canadian Journal of Law & Jurisprudence* 5-36.

Bicchieri, C. et al. 2023. "Social Norms," in E. Zalta and U. Nodelman, eds. *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/social-norms/

Coase, R. 1960. "The Problem of Social Cost," 3 *The Journal of Law and Economics* 1-44.

Coleman, J. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Dari-Mattiacci, G., and Parisi, F. 2006. "The Economics of Tort Law: A Precis," in J. Backhaus, ed. *The Elgar Companion to Law and Economics* (2nd ed.). Cheltenham and Northampton, MA: Edward Elgar Publishing.

De Geest, G. 2013. "N problems require N instruments," 35 *International Review of Law and Economics* 42-57.

Ellickson, R. 1991. *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.

Ellickson, R. 2001. "The Market for Social Norms," 3 *American Law and Economics Review* 1-49.

Elster, J. 2011. "Reciprocity and Norms," in M. Fleurbaey et al., eds. *Social Ethics and Normative Economics. Studies in Choice and Welfare*. Berlin and Heidelberg: Springer.

Epstein, R. 2009. *Simple Rules for a Complex World*. Cambridge and London: Cambridge University Press.

Fon, V., and Parisi, F. 2003. "Reciprocity-Induced Cooperation," 159 *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 76-92.

Fon, V., and Parisi, F. 2007. "On the optimal specificity of legal rules," 3 *Journal of Institutional Economics* 147-164.

Friedman, D. 1995. "Making Sense of English Law Enforcement in the Eighteenth Century," 2 *The University of Chicago Law School Roundtable* 475-505.

Gardner, J. 2008. *Women in Roman Law and Society*. Bloomington and Indianapolis: Indiana University Press.

Glenn, P. 1997. "The Capture, Reconstruction and Marginalization of "Custom"," 45 *American Journal of Comparative Law* 613-620.

Green, J. 2016. *The Persistent Objector Rule in International Law*. Oxford: Oxford University Press.

Guzman, A. 2008. *How International Law Works. A Rational Choice Theory*. New York: Oxford University Press.

Hadfield, G. 2017. *Rules for a Flat World. Why Humans Invented Law and How to Reinvent It for a Complex Global Economy*. New York: Oxford University Press.

Hadfield, G., and Weingast, B. 2012. "What is Law? A Coordination Account of the Characteristics of Legal Order," 4 *The Journal of Legal Analysis* 471-514.

Hathaway, O., and Shapiro, S. 2011. “Outcasting: Enforcement in Domestic and International Law,” 121 *The Yale Law Journal* 252-349.

Hayek, F. 2013. *Law, Legislation, and Liberty. A New Statement of the Liberal Principles of Justice and Political Economy* (1973). London and New York: Routledge.

Hindriks, F., and Guala, F. 2015. “Institutions, Rules, and Equilibria: A Unified Theory,” 11 *Journal of Institutional Economics* 459-80.

Hodgson, G. 2006. “What Are Institutions?,” 40 *Journal of Economic Issues* 1-25.

in‘t Veld, M. 2023. “Was There a Rule of Law in Early Modern Amsterdam? Mercantile Customary Law as a Test,” 15 *Hague Journal of the Rule of Law* 263-282.

Kaplow, L. 1995. “A Model of the Optimal Complexity of Legal Rules,” 11 *Journal of Law, Economics, & Organization* 150-163.

Knight, K. 1992. *Institutions and Social Conflict*. Cambridge: Cambridge University Press.

Koren, G. 1992. “Two-Person Repeated Games Where Players Know Their Own Payoffs,” Courant Institute of Mathematical Sciences.

Leeson, P. 2008. “Social Distance and Self-Enforcing Exchange,” 37 *The Journal of Legal Studies* 161-188.

Leoni, B. 1993. *Freedom and the Law*. Los Angeles: Nash Publishing.

Lintott, A. 1968. *Violence in Republican Rome*. Oxford: Oxford University Press.

Lim, C., and Elias, O. 1997. “The Role of Treaties in the Contemporary International Legal Order,” 66 *Nordic Journal of International Law* 1-21.

Luban, D. 2020. “What Is Spontaneous Order?,” 114 *American Political Science Review* 68-80.

Mahoney, P., and Sanchirico, C. 2001. “Competing Norms and Social Evolution: Is the Fittest Norm Efficient?,” 149 *University of Pennsylvania Law Review* 2027-2062.

McAdams, R. 2000. “Focal Point Theory of Expressive Law,” 86 *Virginia Law Review* 1649-1730.

McAdams, R. 2009. “Beyond the prisoners' dilemma: Coordination, game theory, and law,” 82 *Southern California Law Review* 209-258.

Murray, A. 1983. *Germanic Kinship Structure. Studies in Law and Society in Antiquity and the Early Middle Ages*. Toronto: Pontifical Institute of Mediaeval Studies.

Nguyen, N. 2006. “Roman Rape: An Overview of Roman Rape Laws from the Republican Period to Justinian's Reign,” 13 *Michigan Journal of Gender & Law* 75-112.

Ogus, A. 2002. “The Economic Basis of Legal Culture: Networks and Monopolization,” 22 *Oxford Journal of Legal Studies* 419-434.

Okada, I. et al. 2018. “A solution for private assessment in indirect reciprocity using solitary observation,” 455 *Journal of Theoretical Biology* 7-15.

Parisi, F. 1995. “Toward a theory of spontaneous law,” 6 *Constitutional Political Economy* 211-231.

Parisi, F. 2001. “Sources of Law and the Institutional Design of Lawmaking,” 19 *Journal of Public Finance and Public Choice* 95-122.

Pęski, M. 2014. “Repeated games with incomplete information and discounting,” 9 *Theoretical Economics* 651-694.

Posner, E. 2000. *Law and Social Norms*. Cambridge, MA and London: Harvard University Press.

Postema, G. 1982. “Coordination and Convention at the Foundations of Law,” 11 *The Journal of Legal Studies* 165-203.

Powell, B., and Stringham, E. 2009. “Public Choice and the Economic Analysis of Anarchy: A Survey,” 140 *Public Choice* 503-538.

Raz, J. 1994. *Ethics in the Public Domain. Essays in the Morality of Law and Politics*. Oxford: Oxford University Press.

Robinson, P. 1980. “A Brief History of Distinctions in Criminal Culpability,” 31 *Hastings Law Journal* 815-853.

Rossi, G., and Spagano, S. 2018. “From Custom to Law, An Economic Rationale behind the Black Lettering,” 52 *Journal of Economic Issues* 1109-1124.

Schiller, A. 1938. “Custom in Classical Roman Law,” 24 *Virginia Law Review* 268-282.

Schuck, P. 1992. “Legal Complexity: Some Causes, Consequences, and Cures,” 42 *Duke Law Journal* 1-52.

Shaw, M. 2017. *International Law. Eighth edition*. Cambridge: Cambridge University Press.

Smith, H. 2009. “Community and Custom in Property,” 10 *Theoretical Inquiries in Law* 5-41.

Smith, G. 1922. “Early Greek Codes,” 17 *Classical Philology* 187-201.

Sugden, R. 1986. *The Economics of Rights, Co-operation and Welfare*. New York: Palgrave Macmillan.

Sugden, R. 1989. “Spontaneous Order,” 3 *The Journal of Economic Perspectives* 85-97.

Taylor, M. 1982. *Community, Anarchy, and Liberty*. Cambridge: Cambridge University Press.

Unger, R. 1976. *Law in Modern Society. Toward a Criticism of Social Theory*. New York: The Free Press.

Voigt S. 2024a. “Determinants of social norms I – the role of geography,” 20 *Journal of Institutional Economics* e5.

Voigt, S. 2024b. “Determinants of social norms II – religion and family as mediators,” 20 *Journal of Institutional Economics* e10.

Weber, M. 1949. “’Objectivity’ in social science and social policy,” in: Weber, M., ed. *Essays in the Methodology of the Social Sciences*. New York: The Free Press.

Westbrook, R. 1988. “The Nature and Origins of the Twelve Tables,” 105 *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Romanistische Abteilung* 74-121.

Wormland, J. 1982. “The Blood Feud in Early Modern Scotland,” in: Bossy, J., ed. *Disputes and Settlements. Law and Human Relations in the West*. Cambridge: Cambridge University Press.

Wright, R. 2000. “The Illusion of Simplicity: An Explanation of Why the Law Can't Just Be Less Complex,” 27 *Florida State University Law Review* 715-744.

Young, P. 2001. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

Young, P. 2015. “The Evolution of Social Norms,” 7 *Annual Review of Economics* 359-387.