Zdybel, Karol B.

Working Paper

Norms among heterogeneous agents: a rational-choice model

ILE Working Paper Series, No. 78

Provided in Cooperation with:
University of Hamburg, Institute of Law and Economics (ILE)

Suggested Citation: Zdybel, Karol B. (2024) : Norms among heterogeneous agents: a rational-choice model, ILE Working Paper Series, No. 78, University of Hamburg, Institute of Law and Economics (ILE), Hamburg

This Version is available at:
https://hdl.handle.net/10419/283629

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR RECHTSWISSENSCHAFT

INSTITUTE OF LAW AND ECONOMICS
WORKING PAPER SERIES

Norms among heterogeneous agents: a rational-choice model

Karol B. Zdybel

Working Paper 2023 No. 78

February 2024

Photo by UHH/RRZ/Mentz

Norms among heterogeneous agents: a rational-choice model

Karol B. Zdybel[1]

Abstract:

Spontaneous norms, or simply norms, can be defined as rules of conduct that emerge without intentional design and in the absence of purposeful external coordination. While the law and economics scholarship has formally analyzed spontaneous norms, the analysis has typically been limited to scenarios where agents possess complete information about the interaction structure, including others' understanding of desirable and undesirable outcomes. In contrast, this paper examines spontaneous norms under the assumption of agent heterogeneity and private preferences. By employing a game-theoretical framework, the analysis reveals that norms' lifecycle can be divided into a formative phase and a long-run phase. The formative phase crucially shapes the norm's content and is itself critically dependent on the initial beliefs that agents hold about each other. Moreover, spontaneous norms are resilient to minor shocks to the belief structure but disintegrate when the magnitude of shocks becomes significant. In the final part, the paper highlights the broader implications of its findings, indicating applications in general law and economics, legal anthropology and history, and the sociology of social norms.

Keywords: Spontaneous norms, Social norms, Custom, Private assessment, Legal history

JEL-classification: K00, K10, K39, P48, Z13

[1] European Doctorate in Law and Economics, Institute of Law and Economics, University of Hamburg. Email: zdybel@law.eur.nl. ORCID: https://orcid.org/0000-0002-3432-6311.

The author thanks Michael Faure, Jerg Gutmann, Raphael Maesschalck, Tanja Porčnik, Yvon Rocaboy, Roee Sarel, Norbert Slenzok, and Stefan Voigt for help in improving the paper. All shortcomings are mine.

1. Introduction

Spontaneous norms are typically understood as rules of conduct that emerge without deliberate design and in the absence of purposeful external coordination. Patterns of cooperation may spontaneously develop in repeated interactions. Yet until now, rational choice analysis has predominantly focused on examining conditions that foster the development of cooperative norms in the face of opportunistic motives. The existence of opportunistic motives gives rise to an incentive problem – an issue revolving around incentives to overcome opportunism. Crucially, the notion of cooperation is assumed to be predefined, so that the meaning of terms like “benefit”, “harm”, “defection”, etc., along with the overall structure of the interaction, is known to all agents. However, what if agents possess private understandings of what constitutes desirable and undesirable behavior? This invites an additional coordination problem: a challenge of coordinating efforts to sustain cooperation when the variety of private preferences, coupled with a lack of mutual knowledge, creates a possibility of discoordination.

This paper attempts to understand the formation of spontaneous norms from a rational choice perspective when both incentive and coordination problems are present. Formally, the paper deviates from the conventional assumption of complete information and considers heterogeneous agents with private assessment, i.e., private schemes of preferences regarding desirable behavior. The analysis unfolds with a repeated Bayesian game where players face various “deviation opportunities” that may be collectively rectified. With no certain knowledge of how others assess deviations, the players need to coordinate costly collective action prone to opportunistic motives. We find (à la Bicchieri, 2006) that the initial configuration of beliefs plays a major role in shaping the developmental phase of spontaneous norms, and thus indirectly determines the scope of performances regulated by spontaneous norms in the long run.

After presenting the model, the paper discusses its results. First, it establishes that norms should be stable or even gradually gain complexity when external shocks to cooperation parameters are minor; yet they become unstable with significant shocks. Furthermore, the findings are applied to legal history and legal anthropology in the study of the evolution of early

law. The paper suggests possible sources of differentiation of legal rules applicable within groups (such as clans, tribes, cognatic kinship units, etc.) and outside of such groups – a distinction famously characterized in legal anthropology as a "segmentary lineage system" (Evans-Pritchard, 1940). Moreover, the paper can contribute to the discussion on the tradeoff between social norms and law (c.f., e.g., Ellickson, 1991; McAdams & Rasmusen, 2007; Druzin, 2016; De Geest, 2020), proposing that the efficiency of norms decreases as the complexity of the regulated situation increases.

The article is structured as follows: Section 2 introduces simultaneous incentive and coordination problems in social cooperation and portrays spontaneous norms as one of the methods of resolving them. Section 3 presents a game-theoretical model of spontaneous norms. Section 4 discusses the findings and possible applications. Section 5 offers concluding observations and suggests related research ideas.

2. Norms and social coordination

The majority of economics-inspired research dedicated to norms focuses on the conditions that foster compliance (see, e.g., McAdams & Rasmusen, 2007, for an overview). It recognizes that cooperative norms can organically develop when agents face opportunistic incentives, e.g., incentives to cheat or to avoid the costly punishment of cheaters. Achieving this necessitates a sacrifice of immediate gains in favor of future benefits. For example, in the oft-discussed repeated prisoner's dilemma scenarios, the analysis revolves around identifying conditions under which players consistently choose cooperation over defection.

The corresponding modeling techniques have mirrored the emphasis on compliance by assuming that agents face temptations to defect while possessing ideal knowledge of the game structure (e.g., Ullman-Margalit, 1977; Sugden, 1989; Cooter, 1996; Young, 2001 [1998]; see Bicchieri et al., 2018, for an overview). For instance, agents may be imperfectly monitored or face the last-period problem. However, it is notable that concepts like "cooperation", "defection", "benefit", "harm", etc. are predefined within the game structure which, by

assumption, is known to all. Consequently, the challenge of establishing a shared understanding of desirable and undesirable behavior is absent from the analysis of norms.

However, in realistic scenarios involving heterogeneous agents with less-than-ideal knowledge about each other, this issue becomes crucial. Consider a scenario where a buyer and a seller agree on the purchase of a commodity. The buyer makes the payment and the seller delivers the goods. However, the outcome may deviate from the parties' original expectations: the buyer may find the quality unsatisfactory; the delivery might take longer than expected; the granularity is nonstandard, and so forth. When multiple parties need to jointly expend effort to ensure desirable performance (e.g., when they need to join forces to penalize a wrongdoer, and no third-party enforcement is available) a coordination problem arises. How to recognize events that necessitate this collective effort? Specifically, if deterrence or enforcement of a remedy requires collective action, when should this action be taken?

In other words, establishing norms involves a *dual problem of coordination and compliance*. It combines the challenge of coordinating the actions of multiple agents (i.e., providing a shared understanding of what compliance means) with the challenge of preventing defection (i.e., providing incentives to comply). The connection between coordination and incentive problems has been recently emphasized by several law and economics scholars (e.g., McAdams, 2009; Hadfield & Weingast, 2012; Bertolini, 2016). Hadfield and Weingast state that when agents are heterogeneous,

> "[a]chieving deterrence (…) requires coordinating collective punishment in response to particular actions. This presents two essential problems. First, because each potential punisher has an idiosyncratic logic for assessing wrongfulness, none are able to determine unilaterally when to punish in response to possible rule violations. (…) Second, because punishment is individually costly, punishers need an incentive to punish." (Hadfield & Weingast, 2013, p. 9)

Bertolini outlines the difference between coordination and compliance problems in the following way:

> "Coordination problems arise from the necessity of coordinating individual decisions to punish (…). Assuming that people have incentives to bear the costs of enforcing the norm, coordination problems involve determining how multiple, simultaneous individual decisions to punish can be coordinated to generate a coherent and predictable enforcement process. In comparison, incentive problems arise when self-interested individuals are unwilling (because they have no incentive) to bear the costs of punishing the norm violators." (Bertolini, 2016, p. 16)

2.1. Spontaneous norms as a method of social coordination

While the development of cooperative rules needs to address the dual problem of coordination and compliance, it can be solved in more than one way. Spontaneous norms are one of them that has received considerable attention in the law and economics scholarship. Spontaneous norms are typically construed as complex patterns of cooperative behavior that emerge without deliberate design and purposeful external coordination (cf., Sugden, 1989; Coleman, 1990; Parisi 1995; Young, 2001 [1998]). Extensively examined in both traditional legal theory and law and economics, spontaneous norms and related concepts are viewed through similar lenses by major legal theorists (e.g., Hart, 1994 [1961]; Hayek, 1982 [1976]; Raz, 1994).[2]

[2] For instance, Hart (1994 [1961]) identifies spontaneous norms with primary rules of behavior, i.e., rules regulating conduct that are not accompanied by secondary rules, i.e., rules pertaining to the alteration or interpretation of primary rules. As such, spontaneous norms must be self-evident to those involved. Likewise, Hayek (1982 [1976]) emphasized the importance of emergent-in-fact patterns of conduct that over time become implicitly acceptable as binding rules without the involvement of legal bureaucracy. Raz (1994) considers tradition-oriented norms in contradistinction to a bureaucratic model of the rule of law. According to Raz, while the latter is based on an impartial third-party apparatus that publicly promulgates abstract rules and applies them to individual cases, norms can be identified with proven community practices whose meaning and purpose are tacitly understood by community members.

In the law and economics scholarship, norms have been modeled game-theoretically either as evolutionarily stable equilibria in the evolutionary setting (e.g., Axelrod & Hamilton, 1981; Parisi, 1995; Sugden, 2005 [1986]; Bertolini, 2016; Morsky & Akçay, 2019) or subgame-perfect equilibria in the perfectly rational setting (e.g., Mahoney & Sanchirico, 2003; Bicchieri & Sontuoso, 2020). The conventional law and economics approach posits that the demand for cooperative norms arises in situations of social dilemma, where individual incentives are likely to produce socially inferior outcomes. Spontaneous norms are understood as solutions to social dilemmas crafted entirely by the players within the boundaries of the game (see, e.g., Aoki, 2001; Parisi, 2000). This characteristic distinguishes spontaneous norms from legal rules, conventionally depicted as deliberate changes to the game's structure, wherein reallocations of legal rights mean revisions to the payoff matrix (e.g., Picker, 1994).

Importantly, the textbook approach rules out the assistance of a third party in the creation, clarification, adjusting, or changing of spontaneous norms. In interactions between agents, the meaning of actions is interpreted privately, so that “any normative classification is limited to the classification supplied by individuals acting independently.” (Hadfield & Weingast, 2012, p. 21) Consequently, the coordination problem must be resolved without external involvement; rather, it must be addressed through a confluence of expectations arising from independent decision-making. In the context of the previous example, neither the seller nor the buyer could understand the rules of trade by following third-party statements or reconstructing third-party reasoning.

3. Model

We employ a game-theoretical framework to scrutinize spontaneous norms. A straightforward model, fashioned after the work of Hadfield and Weingast (2012), represents a repeated social dilemma intertwined with a coordination problem. Within this model, agents possess the capacity to engage in cooperative efforts to uphold a norm, yet are confronted with opportunistic incentives. Furthermore, the occurrence of diverse events that may or may

not upset agents' utility introduces an element of uncertainty, as agents lack definitive knowledge regarding the utility effects these events exert on their counterparts.

To offer a more tangible illustration, consider a scenario (similar to the one considered before) in which two buyers repeatedly transact with sellers. When a seller fulfills his end of the transaction, the buyers can enforce corrections to (or compensation for) perceived lapses in the seller's performance, such as late shipment or subpar quality. However, rectifying the seller's performance requires collective action. First of all, no third-party enforcement agency is available to which a wronged buyer could turn for help. Further, we assume that buyers have no effective possibility of single-handedly penalizing misbehaving sellers. For example, a threat of a future boycott by a single buyer could be implausible because buyers transact with the same seller too rarely. Moreover, we assume that a single buyer has too little power, e.g., in terms of physical strength, persuasive capability, or the ability to inflict reputational damage, to force a seller into correcting or compensating for the performance. A potential confrontation would at best result in a one versus one stalemate. Thus, enforcing a correction requires assistance from another enforcer. Two united buyers can overpower a seller, yet not without bearing the cost of a possible conflict.

The choice of collective action as a prerequisite for enforcement is motivated by the potential implications of the analysis. Collective enforcement emerges as a characteristic feature in social environments devoid of centralized mechanisms for establishing social order. This phenomenon is particularly pertinent to historical societies predating the advent of the modern nation-state, as highlighted by multiple scholarly works (Drew, 1995; Friedman, 1995; Allen & Barzel, 2011; Koyama, 2014), notably societies governed by so-called "primitive law" (Hoebel, 1967; Diamond, 1971). Because the analysis is intended to capture certain aspects of the pre-legal and early legal societal organization, it cannot include a centralized enforcement apparatus as a background assumption. Likewise, collective action is one of the few viable enforcement mechanisms in the international system (see, e.g., Bederman, 2001; Guzman, 2008; Shaw, 2017), and plays a key role in the enforcement of social norms (e.g., Axelrod, 1986; Coleman, 1990).

Complicating matters, the buyers may harbor disparate views on which of the sellers' performances are wrongful performances, and they lack certainty about each other's perspectives. This means that the buyers do not know, which of the sellers' actions harm the other buyer and which are neutral. This assumption corresponds to the previously outlined problem of coordination characteristic of all real-world institutions. It reflects an environment consisting of heterogeneous agents who possess less-than-ideal knowledge about each other.

3.1. Formal framework

The game unfolds across an infinite sequence of identical periods (stage games). The inter-period discount factor δ is universally applicable to all players. Therefore, in period T, player's expected utility in the remainder of the game is expressed as $\sum_{t=T}^{\infty} \delta^{t-T} \pi_t$, where π_t signifies the expected utility in period t.

Agents, types, and beliefs. There are two infinitely lived strategic agents identified as Buyer 1 and Buyer 2. These buyers in each period purchase from sellers whose performances are susceptible to imperfections such as late shipment, substandard product quality, reduced quantity, and the like. The set $W = \{w_1, \dots, w_N\}$ includes all conceivable deviations from a flawless performance.

The buyers are heterogeneous in their assessment of received performances. The subset ω_k within W represents the deviations considered wrongful by Buyer k $(= 1, 2)$; the remaining deviations are deemed neutral by this buyer. Consequentially, the subset ω_k delineates Buyer's k type. It encapsulates purchase-related events causing a utility loss or otherwise subjectively perceived as undesirable. In essence, the agent's type reflects a private classification logic: the agent's understanding of rightful and wrongful performances by the seller.

Crucially, buyers' types remain private information, unbeknownst to each other. The buyers do not know each other's types with certainty but hold beliefs about each other. This feature of the framework reflects the idea that encounters between agents are characterized by varying degrees of familiarity with others' preferences, interests, and objectives.

$\{\beta_1^k(t), \ldots, \beta_N^k(t)\}$ signifies Buyer's k beliefs at the outset of period t, where $\beta_i^k(t)$ denotes the probability that Buyer l (at time t) considers deviation w_i wrongful, i.e., the probability that $w_i \in \omega_l$. For the sake of analytical simplicity, the framework presumes that deviations are independent from each other from a buyer's perspective. The assessment of one deviation as harmful imparts no information about the evaluation of another deviation. For example, the fact that one buyer does not tolerate late shipment is no indication of whether this buyer tolerates subpar quality. Formally, for all distinct deviations i and j, $P(w_i \in \omega_l \,\&\, w_j \in \omega_l) = \beta_i^k \beta_j^k$.

Stage game. In every period, both buyers transact with sellers. One of the sellers[3] commits a solitary deviation from flawless performance in each period. The deviation can be envisioned as taking a cost-cutting opportunity: when the opportunity materializes, the seller exploits it and saves a significant part of the cost. Importantly, only one cost-cutting opportunity arises in each period, and it pertains to one buyer exclusively, with equal probability for both. The nature of this deviation is random: p_i stands for the probability of deviation w_i occurring. In one period a seller sends a shipment late, in another period the product is of subpar quality, in yet another period it is of low quantity, etc.

Following the occurrence of the deviation w_i, the buyers observe the seller's performance and simultaneously decide whether to enforce corrective action. Enforcing corrective action is broadly interpreted, encompassing measures such as pressuring the seller into making monetary compensation or compelling specific performance. Critically, enforcement succeeds only if both buyers opt to take action which reflects the model's emphasis on collective punishment as an enforcement mechanism. For instance, both buyers may jointly threaten the seller with adverse consequences unless the quality is immediately improved, with consequences deemed significant enough to motivate the seller to rectify the performance. However, if only one or neither buyer takes action, the seller lacks incentive to address the objection.

[3] This assumption could also read "at most one of the sellers", with self-evident adjustments to the rest of the model and strategies.

In the aftermath of the decisions on whether to react to w_i, the stage game concludes, and the buyers receive payoffs. A buyer who does not find the seller's performance objectionable[4] gains utility G. Conversely, when Buyer k deems the seller's performance wrongful, and it remains uncorrected or uncompensated, the utility from performance diminishes to 0. However, enforcement endeavors come at a cost; for example, they may involve resources for threatening, coercing, or tarnishing the seller's reputation. For simplicity, the model assumes that a buyer bears the cost c of taking action against the seller, irrespective of whether the other buyer participates. Furthermore, $G > 2c$ is assumed and g is defined as $\frac{G}{2}$ for simplicity. One interpretation of this assumption is that the social gain from collective punishment by two agents exceeds the social cost. Alternatively, it suggests that without taking discounting into account, punishing twice and enforcing correction to one seller's misperformance constitutes a net utility gain.

Solution concept and equilibrium outline. At any point in the aforementioned supergame, buyers strive to maximize their expected payoffs $\sum_{t=T}^{\infty} \delta^{t-T} \pi_t$ while facing the dual problem of coordination and compliance. First, they need to determine the circumstances warranting collective action. In other words, they need a coordination method. Furthermore, even when such circumstances are determined, the persistent temptation to defect and ignore the seller's deviation looms. This arises from the fact that only one buyer can be wronged by the seller's performance in a single period. Consequently, the non-affected buyer can opportunistically save the cost of enforcement that he is otherwise required to incur, as defection entails no immediate utility loss. To preclude such scenarios, agents must address the compliance problem by fostering incentives against defection.

As previously hinted, the emergence of spontaneous norms serves as a potential solution to the dual problem of interpretation and compliance. It was noted that spontaneous norms can be defined as patterns of cooperation that evolve and persist without deliberate coordination by a third party. This implies that a shared understanding of permissible and

[4] Which happens either when $w_i \notin \omega_k$, i.e., when the buyer does not care about w_i, or w_i has been corrected.

unpermissible behavior must organically develop within the group of cooperating agents. To use the context of the previous example, the buyers need to establish cases in which they would jointly penalize deviating sellers. In line with this notion, the forthcoming presentation introduces a perfect Bayesian equilibrium wherein buyers gradually develop a shared understanding of punishable performances and collectively punish them in the long run.

As will be seen, the initial configuration of beliefs $\{\beta_1^1(1), \dots, \beta_N^1(1)\}$, $\{\beta_1^2(1), \dots, \beta_N^2(1)\}$ will play a pivotal role in shaping long-term equilibria in which buyers cooperatively punish deviations. The equilibrium idea is straightforward: buyers anticipate reciprocal cooperation in penalizing deviation w_i if *both* possess sufficiently strong convictions that *both* consider w_i wrongful. Two-way beliefs support underpin endeavors to establish reciprocity-based cooperation.

Moreover, cooperative norms may occasionally develop without an initial expectation of reciprocity. This can happen in an asymmetrical case when only one buyer is confident that he shares an interest in punishing w_i with the other buyer. The better-informed buyer initiates punishment for w_i to signal that he considers w_i wrongful. As a consequence, the expectation of reciprocity for the future is established. This second situation of unilateral initiation of cooperation will be termed norm entrepreneurship.

3.2. Expectation of reciprocity

We will now scrutinize the conditions prompting a buyer to engage in reciprocity-based cooperation when penalizing the seller's performance w_i. Suppose that a seller seized the cost-cutting opportunity w_i against Buyer l in the current period. Moreover, suppose that Buyer k entertains the following expectation that can be dubbed the *expectation of reciprocity*: if Buyer l considers w_i wrongful, he would punish in the current period. Moreover, he would continue punishing any seller who performs w_i toward any of the buyers in the future as long as Buyer k keeps doing the same. Otherwise, Buyer l would never again punish a seller who performs w_i. Given this expectation, Buyer k is incentivized to punish the seller in the current period if the following condition is met:

$$-(1-\delta_i)c + \beta_i^k \delta_i (g-c) > 0 \quad (1)$$

where

$$\delta_i = \delta_i(\delta, p_i) = \frac{\delta p_i}{\delta p_i + 1 - \delta} \quad (2)$$

is a factor discounting every two periods in which the same deviation w_i is expected to occur.[5]

The left-hand side of Inequality (1) represents the expected value of punishing the seller who seized the cost-cutting opportunity w_i. The buyer must incur the cost of punishment c in the present period. Should the buyer choose to punish, both will continue collectively punishing every w_i provided that Buyer l privately considers w_i wrongful – which Buyer k believes to be the case with a probability β_i^k. This means an expected per-period payoff $\beta_i^k(g-c)$ in all future periods in which w_i is comitted. The future stream of utility from cooperation in punishing sellers who commit w_i is proportional to the belief that the preferences about w_i are shared.

Conversely, if Buyer l does not share the negative assessment of w_i, he will take no action against the seller, irrespective of Buyer's k actions: no utility is thus expected in the future with probability $1-\beta_i^k$. The relative weights of utility derived in the current period versus the aggregate utility obtained in all future periods in which sellers are anticipated to perform w_i are $1-\delta_i$ and δ_i, respectively.

In turn, ignoring the deviation w_i means that reciprocity-based cooperation would not be established. Buyer l begins to ignore all future occurrences of w_i and no utility is expected as a result of collective punishment. This is reflected in the right-hand side of Expression (1). With the use of (2), Condition (1) can be simplified to:

$$-(1-\delta)c + \beta_i^k p_i \delta (g-c) > 0 \quad (3)$$

(3) yields the following limiting condition for β_i^k:

$$\beta_i^k > \frac{1-\delta}{\delta} \frac{c}{p_i(g-c)} \equiv \beta^* \quad (4)$$

[5] The probability that it would take exactly K periods before the same opportunity is taken again is given by $p_i(1-p_i)^{K-1}$. Therefore, the expected discount factor $\delta_i(\delta, p_i)$ between two periods in which the seller exploits the same deviation is $\delta_i(\delta, p_i) = \sum_{K=1}^{\infty} \delta^K p_i (1-p_i)^{K-1}$ which after simplification gives expression (2).

which specifies Buyer's k minimum belief β_i^k that is sufficient to engage in cooperation in punishing w_i, provided that Buyer k entertains the expectation of reciprocity.

Condition (4) depends on several factors. First, its stringency heightens with an increase in the number of possible deviations N, or the augmented "complexity" of the interaction between sellers and buyers. An increased N translates to a diminished probability p_i of the deviation w_i occurring in any given period. When the performance of a seller is more complex (e.g., involving a product or service made of multiple, disparate, and intricate items that are potentially prone to various flaws or failures), heterogeneous agents need a more robust belief that others share their assessment of undesirable performances before venturing into reciprocal cooperation.

Naturally, the stringency of Condition (4) eases under an improved cost-benefit balance, evident when the gain g from proper performance increases, or the cost of punishment c decreases. Conversely, the condition tightens when this balance deteriorates. Lastly, agents' inclination toward cooperation is heightened as they place a higher value on future utility relative to current utility. Consequently, Condition (4) becomes more lenient with an elevated discount factor δ. In short,

$$\frac{\partial \beta^*}{\partial p_i} < 0; \frac{\partial \beta^*}{\partial g} < 0; \frac{\partial \beta^*}{\partial c} > 0; \frac{\partial \beta^*}{\partial \delta} < 0. \quad (5)$$

Crucially, Condition (4) delineates the necessary incentives for a buyer to penalize a seller who exploited the opportunity w_i, contingent on the expectation of reciprocity. This qualification is essential: it means that the buyer is incentivized to punish a seller if he believes that Condition (4) is satisfied for the other buyer as well. It is only when Condition (4) is fulfilled for both buyers, i.e., $\beta_i^1 > \beta^*$ and $\beta_i^2 > \beta^*$, that they would find it advantageous to collectively penalize the seller in the current period. In the event that Condition (4) is met for only one buyer and not the other, the latter would refrain from punishing the seller in the current period, thereby undermining the feasibility of the expectation of reciprocity.

3.3. Norm entrepreneurship

Up to this point, it has been asserted that robust two-sided beliefs that buyers' private preferences regarding w_i are aligned pave the way for reciprocity-based cooperation. However, the potential for establishing long-term cooperation is not yet exhausted.

Consider a scenario where Condition (4) is not satisfied for both buyers. As stated, reciprocity-based cooperation will neither begin nor continue. Yet, there exists an avenue for one buyer to rectify this situation. Specifically, suppose a buyer signals their disapproval of performance w_i by unilaterally imposing punishment on the seller responsible for committing this performance. Additionally, assume the following: after the period in which w_i is unilaterally penalized by Buyer k, Buyer's l belief β_i^l is updated to 1, Condition (4) is satisfied for both buyers, and both buyers begin to harbor the expectation of reciprocity from this point onward. The meaning of this assumption is straightforward. By signaling a disapproval of w_i, a buyer informs the other agent about the possibility of future cooperation in punishing sellers who seize cost-cutting opportunities w_i. If the signaling agent already believes in this possibility, the expectation of reciprocity becomes mutual.

If the assumptions above hold, unilateral punishment is indeed a signal: using costly punishment to demonstrate that a buyer considers the seller's performance w_i wrongful is less costly for someone who genuinely considers it wrongful. This is because, for a truthful agent, the cost is offset by the potential benefits derived from future cooperation in penalizing sellers exploiting w_i. Thus, the initiation of unilateral punishment conveys information about the buyer's type. The incentive to signal is in place if:

$$-c + \beta_i^k \delta_i \sum_{r=0}^{\infty} {\delta_i}^r (g - c) - (1 - \beta_i^k)\delta_i c > 0 \tag{6}$$

which is equivalent to:

$$-(1 - \delta_i)c + \beta_i^k \delta_i (g - c) - \left(1 - \beta_i^k\right)\delta_i (1 - \delta_i)c > 0 \tag{7}$$

The left-hand side of Expression (7) represents the Buyer's k expected payoff if he signals his type by unilaterally punishing the seller who exploited the cost-cutting opportunity w_i. Its interpretation can be summarized as follows. In the current period, Buyer k incurs the

cost of punishment c. The subsequent period in which the cost-cutting opportunity w_i is expected to occur is discounted with the discount factor δ_i. During this period, beliefs are updated and the expectation of reciprocity exists between the buyers. Thus, Buyer k believes that with probability β_i^k, both buyers would begin collectively punishing any seller who exploits w_i throughout the remainder of the game after the beliefs are updated. In this case, Buyer's k expected per-period utility amounts to $g - c$.

However, Buyer k believes that with probability $1 - \beta_i^k$, private preferences regarding w_i are not aligned and thus the future attempt to cooperate would fail. In this scenario, Buyer k incurs the cost c again, but Buyer l does not reciprocate. Finally, the right-hand side of Expression (7) represents the expected utility of ignoring w_i in the current period. Beliefs are never updated and buyers never engage in collective punishment of sellers committing deviation w_i.

The concept underpinning Condition (7) can be termed *norm entrepreneurship*. Norm entrepreneurship is an idea originating from the law and economics scholarship; it suggests that changes to prevailing practices can be initiated by a change agent. The change agent, or norm entrepreneur, is an individual who disrupts preexisting social expectations and attempts to usher in a new mode of conduct (see, Ellickson, 2001; Bicchieri, 2016). If successful, a norm entrepreneur triggers a process of adjustments in social behavior. For instance, a norm entrepreneur might be the first to object to the typically accepted practice of late delivery. Despite the risk of repercussions such as loss of good reputation among the sellers or other adverse consequences, this individual's actions may inspire others to follow suit, ultimately leading to a change in group behavior.

Therefore, Expression (7) specifies the incentive necessary to engage in norm entrepreneurship. It is equivalent to a limiting condition imposed on Buyer's k belief β_i^k. After substituting δ_i, this limiting condition takes the following form:

$$\beta_i^k > \frac{1-\delta}{\delta}\frac{c}{p_i}\frac{1-\delta+2\delta p_i}{(1-\delta)g+\delta p_i(g-c)} \equiv \beta^{**} \tag{8}$$

It can be verified that $\beta^{**} > \beta^{*}$ holds as long as $\delta < 1$ and $\beta^{*} \leq 1$ are satisfied. Thus, norm entrepreneurship necessitates a more robust belief in the shared assessment of sellers' performance w_i compared to reciprocity-based cooperation.

The rationale is straightforward: the introduction of a signaling phase postpones the prospect of long-term cooperation. To offset this delay, a higher level of confidence is required that cooperation in penalizing w_i will ultimately materialize. This aligns with the results from the literature on norm entrepreneurship, suggesting that norm entrepreneurship involves a risk of instigating social change without certainty of success. Therefore, successful norm entrepreneurs turn out to be agents who enjoy a comparative advantage in bearing such risk, such as those with a greater interest in changing norms, the capacity to accommodate the associated costs, or superior knowledge of the relevant circumstances (see, e.g., Cooter, 1996; Mackie, 1996; Ellickson, 2001; Acemoglu & Jackson, 2014). In the context of the model, superior knowledge can be interpreted as particularly high values of β_i^k, i.e., knowledge of others' private preferences.

However, despite superficial similarities, a crucial distinction exists between Inequality (4), which sufficed for establishing reciprocity-based cooperation, and condition (8). Unlike Condition (4), Condition (8) is asymmetrical: it needs to be satisfied only for a single buyer to incentivize a cooperation attempt. This asymmetry reinforces the analogy with norm entrepreneurship. Those holding particularly strong beliefs that initiating behavioral changes may eventually succeed are more likely to initiate them; others are in a position to become late adopters.

Finally, Condition (8) depends on the complexity of seller's performance N, gains from due performance, the cost of engaging in punishment c, and the discount factor δ, akin to the manner in which Condition (4) operates. In other words:

$$\frac{\partial \beta^{**}}{\partial N} > 0; \frac{\partial \beta^{**}}{\partial c} > 0; \frac{\partial \beta^{**}}{\partial g} < 0; 0 \frac{\partial \beta^{**}}{\partial \delta} < 0. \quad (9)$$

3.4. Equilibrium

In light of the preceding considerations, it becomes possible to characterize the equilibrium. Equilibrium behavior manifests in a relatively straightforward manner, comprising two phases: a formative phase and a long run. In the formative phase – i.e., in the first periods in which deviations w_i ever occur – the buyers delineate the events that trigger collective punishment by engaging either in reciprocity-based cooperation or norm entrepreneurship. If neither avenue is viable, the deviation is permanently ignored. After the completion of the formative phase, the scope of cooperation becomes fixed and persists indefinitely from this point onward; this constitutes the long-run phase.

As previously suggested, the initial configuration of beliefs plays a decisive role in shaping the scope of cooperation. In scenarios where agents are heterogeneous and possess incomplete information about each other, the structure of preexisting mutual knowledge dictates the trajectory of the formative phase. This episode, in turn, determines the array of deviations w_i that agents systematically penalize in a coordinated manner in the long run. To illustrate this relationship between beliefs and equilibrium, a $1x1$ two-dimensional belief space is depicted in Figure 1 below.

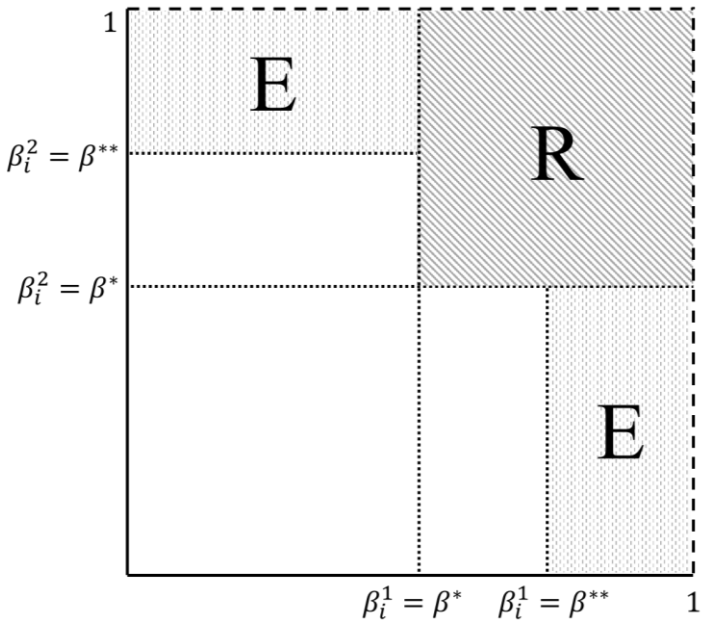

Figure 1. Buyers' initial beliefs and corresponding equilibrium behavior

The belief space can be partitioned into three distinct areas: R, E, and the remaining segment. Area R encompasses all combinations of beliefs β_i^1, β_i^2 such that both buyers believe that they share a negative assessment of sellers' behavior w_i with a probability at least β^* (i.e., $\beta_i^1 > \beta^*$ and $\beta_i^2 > \beta^*$). Essentially, initial belief combinations within area R support mutual expectations of reciprocity, thus sufficiently incentivizing reciprocity-based cooperation.

Conversely, area E comprises asymmetrical combinations of beliefs that instigate norm entrepreneurship. Buyer k, if interested in consistently correcting performance w_i in the future, strongly presupposes a similar interest from the other buyer (i.e., $\beta_i^k > \beta^{**}$). However, this buyer also recognizes that the other buyer lacks the necessary information for immediately entering reciprocity-based cooperation (i.e., $\beta_i^l < \beta^*$). Consequently, Buyer k penalizes w_i without immediate expectations of success; rather, the goal is to update Buyer's l beliefs, allowing the expectation of reciprocity in penalizing w_i to materialize in the future. Hence, when w_i occurs, and buyers' beliefs reside in area E, punishment is expected from the better-informed Buyer k whose belief $\beta_i^k > \beta^{**}$.

In summary, we say that punishment after a seller performed w_i is expected from a buyer if either the buyers' beliefs β_i^1, β_i^2 are in R or when their beliefs are in E, and the buyer is the better-informed agent.

With this in mind, the equilibrium strategies and beliefs can be specified as follows:

Buyer's k $(= 1,2)$ strategy:

1) If deviation w_i occurred for the first time, $w_i \in \omega_k$, and Buyer k is expected to punish w_i, punish; otherwise, ignore w_i;
2) If deviation w_i occurred at least once before, $w_i \in \omega_k$, Buyer k is expected to punish w_i, and no buyer who was expected to punish w_i in the past failed to do so, punish; otherwise, ignore w_i.

Buyer's k $(= 1,2)$ beliefs $\beta_i^k(t)$ for $i = 1, \dots, N$:

1) $\beta_i^k(t) = 1$ if Buyer l always succeeded in punishing w_i when expected to do so until period $t - 1$;

2) $\beta_l^k(t) = 0$ if Buyer l failed to punish w_i at least once when expected to do so until period $t-1$;

3) $\beta_l^k(t) = \beta_l^k(1)$ otherwise.

Within the first two periods when each deviation w_i occurs, i.e., in the formative phase, buyers endeavor to establish cooperation for the collective punishment of of w_i. This pursuit unfolds through either reciprocity-based cooperation or norm entrepreneurship. In the former scenario, both buyers penalize the seller who performed w_i, anticipating a reciprocal response from the other buyer. In the latter scenario, a better-informed buyer initiates punishment, hoping for the other one to follow in the future.

Buyers who are expected to punish deviation w_i and successfully do so reveal that they are interested in penalizing sellers who would commit w_i in the future. This prompts their peers to update the relevant belief to 1, reinforcing the prerequisites for collective punishment. Once started, coordinated punishment following a specific deviation w_i persists indefinitely, as $\beta_l^1 = \beta_l^2 = 1$. On the other hand, buyers expected to punish deviation w_i but failing to do so reveals that they do not consider this deviation wrongful. This failure prompts their peers to update the relevant belief to 0; conditions necessary for cooperation are nullified. As a consequence of belief adjustments in the formative phase, buyers' beliefs stabilize, and so does their behavior. In the long term, buyers systematically penalize sellers deviating from flawless performance in ways deemed harmful by both, provided that the initial beliefs about buyers' private assessment of those deviations resided within areas R or E.

Possible extensions: quid pro quo. It is essential to highlight that the logic of norm entrepreneurship may be used to extend the long-run equilibrium. Beyond cooperation founded on reciprocity, where buyers punish performances that they jointly consider undesirable, *quid pro quo* cooperation may emerge. In *quid pro quo* cooperation, buyers "trade" their assistance in penalizing deviations deemed wrongful by one party only. As a result, buyers might occasionally penalize deviations they privately consider harmless in return for others' assistance in penalizing deviations they, but not the others, consider harmful. The identification

of performances suitable for such trade would require even lengthier procedures compared to previously characterized norm entrepreneurship. For example, it may entail persistent unilateral punishment of sellers committing $w_i \in \omega_k$ by Buyer k (when Buyer k believes that $w_i \notin \omega_l$) until Buyer l begins to likewise unilaterally penalize another deviation $w_j \in \omega_l$ (such that Buyer l believes that $w_j \notin \omega_k$). At this point, the buyers exchange information that is sufficient for mutual assistance in the collective punishment of both w_i and w_j from that moment onward.

The addition of *quid pro quo* cooperation would broaden the scope of deviations that become collectively punished in the long-run equilibrium. However, the conditions for using the procedures for establishing such *quid pro quo* are stringent. Similarly to norm entrepreneurship outlined in the preceding part of this section, these procedures would rely on two-way communication of agents' private preferences through costly actions, initiated by a single agent. However, due to increased complexity, particularly high values of beliefs β_i^k would be necessary to meet the participation constraint. Agents would not risk lengthy and costly attempts to initiate *quid pro quo* cooperation unless they strongly believe that these attempts will be greatly rewarded in the future.

4. Discussion

The foregoing section provided a formal analysis of spontaneous norms emerging among heterogeneous agents with less-than-perfect knowledge about each other. The current section will delve into the findings. Firstly, it will offer a general and non-technical interpretation, situating spontaneous norms within the broader landscape of law and economics research and general legal theory. Secondly, the section will discuss the robustness of spontaneous norms, i.e., the degree to which they can be detailed and specific. In the next step, the discussion will address dynamic stability and the role of the time factor. We posit that, all else being equal, once emerged norms are stable. Moreover, we assert that if the agents' beliefs undergo occasional random shocks of minor magnitude, norms should progressively become more

detailed and specific over time. On the other hand, major shocks undermine norm stability. Finally, we propose the existence of an inverse relationship between the robustness (or complexity) of spontaneous norms and group cohesion. If substantiated, this analysis could illuminate the economic foundations of societies predominantly governed by traditional, gradually developed, or unwritten rules of behavior (commonly termed “customary”), as depicted in classical legal anthropology. Specifically, it may elucidate why such societies often display group identity-based internal organization and why applicable norms vary depending on the involved parties’ group identities.

4.1. General discussion

The coordinative role played by rules received attention from law and economics scholars, yet almost exclusively in the context of theorizing legal orders. Works by Hadfield and Weingast (2012; 2013; 2015 with Carugati) and McAdams (2000; 2004 with Ginsburg; 2005 with Nadler; 2009) in two series of related papers have presented compelling accounts of legal orders as coordination devices. Third-party coordination provided by the law and its institutional representatives (e.g., lawmakers, courts, and other specialized officials) has been portrayed as a viable solution to the dual problem of coordination and compliance. When agents face multiple possibilities to establish cooperation (a game with multiple equilibria) or uncertainty about agents’ private preferences (incomplete information), a convention delegating equilibrium selection to a specialized agent may play a facilitating role. The above-mentioned models mirror the Hartian account of law that emphasizes the division of labor in identifying, altering, applying, and enforcing the law as a prerequisite to an advanced legal system (Hart, 1994 [1961]; Postema, 1982; Lefkowitz, 2017).

The current paper extends the previous theoretical accounts by considering modes of cooperation beyond legal orders based on organized legal bureaucracy. Recognizing that the dual problem of coordination and compliance is inherent in all cooperation attempts (and thus must be solved in all cooperative endeavors), the paper models a scenario in which agents organically develop a shared definition of undesirable behavior and engage in enforcing it, without relying on the assistance of third parties like lawmakers or judges. Instead, in our

interpretation, spontaneous norms undergo trial-and-error procedures during the formative phase before solidifying into a stable pattern of group behavior. Technically, this concept aligns with a well-known equilibrium design for infinite games with incomplete information. In this type of equilibrium play, players reveal private information within a finite number of periods, and the remainder of the game is played without beliefs being updated any further (cf., Koren, 1992; Pęski, 2014). The resemblance between this generic game-theoretical concept and the approach assumed in this paper seems self-evident.

While trial-and-error representations of norm emergence have been present in models within evolutionary game theory (e.g., Sugden, 1986; Young, 2001 [1998]; Aoki, 2001), our model is consistent with the rational choice perspective. Moreover, despite its elementary nature, the model modestly contributes to the growing literature on cooperative norms that emerge under conditions of "private assessment", wherein agents have a private understanding of benefit and harm (see, Okada, 2020 for an overview). Until now, the scholarship on private assessment has been predominantly embedded in theoretical biology rather than law and economics.

It should be emphasized that our approach suggests that spontaneous norms combine the shared practice of systematically penalizing certain behaviors with a shared evaluation of those behaviors. Such norms differ from mere private opinions and actions based on those opinions. Unlike the former, spontaneous norms develop as common standards of behavior that, once established, guarantee that members of the general public would systematically enforce them and abide by them. Thus, the model supports the notion of norms as regularities in group behavior, such that each agent "must view the regularity as a common, public standard, as opposed to seeing it as just a rule for me" (Postema, 2012, p. 716). Concomitantly, as suggested earlier, the distinctive characteristic of spontaneous norms lies in the lack of third-party assistance in the creation, adjustment, or interpretation of such standards.

The preceding section has also identified two fundamental mechanisms through which spontaneous norms originate and persist: reciprocity and norm entrepreneurship. Until now, scholars within the rational-choice framework and those studying the evolution of norms have

consistently recognized reciprocity as a key factor in facilitating cooperation (see, e.g., Axelrod & Hamilton, 1981; Boyd & Richerson, 1988; Fon & Parisi, 2003; Okada et al., 2018). More interestingly, we highlight the role of norm entrepreneurship, i.e., deliberate risk-taking aimed at changing the prevailing patterns of group behavior (Ellickson, 2001). Our model implies that when group members possess symmetrical preexisting knowledge about each other (i.e., when $\beta_i^1 \approx \beta_i^2$ for $i = 1, \dots, N$ at the outset of the game), the expectation of reciprocity plays a more significant role in shaping norms. In this case, agents' initial beliefs are sufficient to directly prompt reciprocity-based cooperation. On the other hand, in scenarios involving asymmetrical knowledge (e.g., when some agents intellectually specialize in some fields or contacts with selected groups of agents) norm entrepreneurship becomes a favored mechanism for norm creation. In such a setting, better-informed agents take the role of pioneers in shaping the patterns of group behavior. Reciprocity-based cooperation develops later, as more agents adopt the norm.

4.2. Robustness of norms

Cooperation based on spontaneous norms may exhibit various degrees of robustness. Spontaneously emergent rules can extensively regulate behavior, often imposing punishments and rewards. Alternatively, they may be "thin" in the sense of governing only a few interactions, leaving the rest to free-for-all, no-holds-barred behavior.

Within the framework developed in the preceding section, the robustness of spontaneous norms can be characterized as the proportion of deviations w_i that become collectively punishable in the long run. Making this metric meaningful requires two further assumptions. First, it assumes that the share of deviations w_i considered wrongful by *both* agents remains constant when other parameters change.[6] Otherwise, robustness would be affected by the varying degree to which agents' private preferences overlap. Second, it assumes that pairs of beliefs β_i^1, β_i^2 for $i = 1, \dots, N$ come from a single probability distribution that is independent from other model parameters. This means that changing model parameters

[6] Without loss of generality, this share may be assumed to be 1.

do not affect the probability that a pair β_i^1, β_i^2 would reside in any given subspace of the belief space.

In light of the aforementioned assumptions, the analysis of the robustness of spontaneous norms naturally follows. The robustness may be gauged by assessing the total area of R and E in Figure 1.[7] The rationale is straightforward. A pair of initial beliefs β_i^1, β_i^2, treated as a realization of a random variable taken from a distribution common to all i, would fall more likely within either R or E when the sum of those areas is greater. In turn, this would imply a greater probability of initial conditions that are sufficient to establish cooperation in penalizing a specific deviation.

The derivates summarized in the groups of inequalities (5) and (9) indicate that the constraints of area R as well as E in Figure 1 (i.e., β^* and β^{**}, respectively) are contingent on several factors. Specifically, they become stricter as the utility derived from due performance diminishes, the costs associated with implementing punishment increase, the discount factor diminishes and, most importantly, as the interaction governed by spontaneous norms becomes more complex.[8] When sellers' performance becomes sufficiently complex (i.e., N is sufficiently large), initial conditions necessary for the entire formative phase sharpen or may even cease to exist. Consequently, only those misconducts that occur frequently become collectively penalized. Less frequent transgressions are prone to indefinitely retaining their status as

[7] This method naturally extends to cases in which individual deviations w_i are characterized by specific probabilities of occurence p_i. Then, the relevant figure is the sum of areas R and E calculated individually for each i, divided by the total area of N individual belief spaces.

[8] Complexity of norms has attracted considerable attention from law and economics scholars in the past several dacades. Moreover, it has been approached and interpreted in diverse manners. One prominent perspective frames the issue of complexity as a choice between "rules" and "standards" (e.g., Kaplow, 1992; Fon & Parisi, 2007) representing opposite poles of legal precision. In this sense, a rule entails "an advance determination of what conduct is permissible, leaving only factual issues for the adjudicator" (e.g., mandating that construction workers wear safety helmets). On the other hand, a standard leaves "both specification of what conduct is permissible and factual issues for the adjudicator" (Kaplow, 1992, p. 559-560) (e.g., requiring that construction workers be adequately protected agains injuries).
Alternatively, complexity has been be understood as the degree to which rules are case-specific. General rules would be applicable across a broad spectrum of situations, regardless while specific rules would be tailor-made for specific circumstances (Kaplow, 1995; Mahoney & Sanchirico, 2005). Our model presented in Section 3 explicitly addresses complexity in this sense. As explained earlier, the parameter N represents the number of potential identifiable deviations committed by sellers (such as late shipment, subpar quality, non-standard glanularity, etc.) that may be perceived as wrongful by the buyers. Hence, a higher N translates to a more complex interaction.

inconsequential actions that fail to elicit a collective response, even if, taken collectively, they constitute the majority of harm inflicted among agents. Put differently, only very high levels of preexisting familiarity between the buyers, as measured by $\beta_i^k(1)$ for $i = 1, \dots, N$, may lead to the emergence of norms regulating such complex conduct.

4.3. Stability and the role of time

The exposition of spontaneous norms has concluded by dividing their dynamics into two distinct phases: the formative phase and long-term equilibrium. It has been stated that the former features a volatile process of trial and error, where agents engage in cooperative attempts and assume risks to usher new patterns of group behavior. In the latter phase, beliefs and actions stabilize, leading to a consistent and replicable form of cooperation.

We can now delve into the dynamic stability of spontaneous norms. To proceed formally, we assume that agents' beliefs in any period t undergo minor random perturbations. In realistic scenarios, agents' beliefs may be perturbed for several reasons. For instance, random and unpredictable events (like social disruptions or technological shocks) may disturb agents' perception of others or intergenerational transmission of culture, including beliefs, may be imperfect. Consequently, Agent's k actual belief $\tilde{\beta}_i^k(t)$ will be defined as $\beta_i^k(t) + \epsilon(t)$ for $i = 1, \dots, N$, where ϵ is random, $E(\epsilon) = 0$, and its absolute value is limited and small; the maximum absolute value of ϵ will be denoted $\bar{\epsilon}$. Importantly, we treat the random perturbation ϵ as a one-off event that does not continue into the future. Actions taken in period t are based on perturbed belief $\tilde{\beta}_i^k(t)$ but result in updating only the basic belief $\beta_i^k(t+1)$. Subsequently, in the next period, the belief relevant for the agent's choice of action becomes $\beta_i^k(t+1) + \epsilon(t+1)$, influencing the choice of action in period $t+1$, and so forth.[9] Moreover, when $\beta_i^k(t) + \epsilon(t)$ exceeds the limiting values of 0 or 1, $\tilde{\beta}_i^k(t)$ takes on this extreme value instead. Finally, for the sake of simplicity, we assume that agents do not observe perturbances ϵ of others' beliefs nor are aware of them.

[9] Technically, the random component added to the beliefs at time t is akin to white noise, not to random walk.

The conclusions derived from the above-mentioned amendments to the model are relatively straightforward. Occasional minor perturbations in beliefs should result in a unidirectional drift toward more robust norms. The rationale is simple. Established patterns of cooperation in penalizing deviations w_i remain resilient to minor shocks as long as the following condition holds:

$$\beta^* < 1 - \bar{\epsilon} \tag{10}$$

Under this assumption, even when a random perturbation ϵ in any given period achieves the value $-\bar{\epsilon}$, Condition (4) is still satisfied. Since agents who already cooperate in collectively penalizing w_i hold beliefs $\beta_i^k = 1$, the condition sufficient for ongoing cooperation cannot be invalidated even after a slight disturbance in beliefs. In other words, under Assumption (10), agents' shared understanding of performances that constitute wrongdoings endure.

Moreover, perturbations of beliefs may sporadically trigger norm entrepreneurship; in turn, successful norm entrepreneurship will increase the robustness of norms. This is because as long as $\beta_i^k(1) > \beta^{**} - \bar{\epsilon}$, each realization of perturbation ϵ implies a nonzero probability that $\tilde{\beta}_i^k(t)$ will exceed the threshold β^{**}. Upon surpassing this threshold, Condition (8) is satisfied for at least one of the agents, incentivizing norm entrepreneurship. Importantly, the probability of a sufficiently large perturbation ϵ occuring before or in period T approaches 1 as T increases to infinity. Consequently, as time progresses, more entrepreneurship attempts will be randomly triggered and spontaneous norms will become increasingly robust.

Conversely, the opposite occurs when the magnitude of transpiring shocks is significant – more precisely, when $\bar{\epsilon}$ is high enough to invalidate Condition (10). In such instances, external shocks will occasionally reach the magnitude that voids Condition (4) – i.e., the condition necessary for reciprocity-based cooperation. This would cause a disruption in cooperation and prompt an update of the relevant belief to 0. Cooperation cannot be subsequently restored without a new formative phase being put in motion. Thus, major external shocks may launch a process of decomposition of a preexisting norm.

In summary, it can be concluded that, unless the underlying conditions undergo major external shocks, spontaneous norms should exhibit stability in the long term. Moreover, in the presence of minor random disturbances (like minor one-time shifts in beliefs), they may slowly become more robust over time, even after the completion of the formative phase. On the other hand, major external shocks are expected to disrupt long-term equilibrium, derailing established norms. In such cases, cooperation can be restored only through a repetition of the norm-building process. This result resembles the “tipping”, and “punctuated equilibrium” dynamics of social norms found with the tools of evolutionary game theory (Young, 2015), where norms are characterized by “long periods of no change punctuated by occasional bursts of activity in which an old norm is rapidly displaced by a new one” (p. 363).

4.4. Group cohesion

The discussion can now progress to an examination of the relationship between the robustness of norms and group cohesion. Theoretical insights from the realms of law and economics suggest that the establishment of private ordering based on group norms requires close-knit and multifaceted relationships among community members (Taylor, 1982; Ellickson, 1991; Cronin, 1999; Bertolini, 2016). These relationships serve a dual role in the decentralized development and enforcement of norms. Firstly, they enhance the flow of information, thereby facilitating the identification of misbehaving agents. Furthermore, they amplify the severity of sanctions arising from a breach of such relationships. The converse proposition, i.e., that lawless environments tend to *produce* communities characterized by these attributes, is less frequently asserted (e.g., Greif & Tabellini, 2017).

Our contribution to this discussion involves a more detailed elaboration on the relationship between the robustness of rules and group cohesion. The model, wherein agents possess less-than-perfect knowledge about each other, allows for the highlighting of this relationship. As previously stated, the model from Section 3 suggests that spontaneous norms are expected to be less robust when the interactions governed by such norms become more complex. The extension of this proposition suggests that more robust norms thrive in more cohesive social environments; as groups become less cohesive, the norms lose robustness

and progressively become less sophisticated. This claim can contribute to elucidating the utility of collectivity-based social structures (like cognatic groups, clans, tribes, lineages, and similar structures) characteristic of societies governed by "primitive" or archaic law, in which norms frequently originate from traditions and remain unwritten, with distinct legal officials being either weak or entirely absent (see, Evans-Pritchard, 1940; Hoebel, 1967; Diamond, 1971).

In this context, a straightforward assertion emerges: group structures delineate the boundaries in which norms characterized by given levels of robustness can be upheld. This delineation can be achieved by either naturally or artificially defining groups in a specific way: namely, that a shared group identity implies a belief that a specific threshold of common interest is met. For example, members within a closely-knit group, such as an extended family, assume a higher degree of alignment in their private preferences. Conversely, members of more socially distant groups presume a more modest level of alignment, and this pattern continues across varying degrees of group proximity.

A natural inference follows: more cohesive groups will be inclined to foster more robust spontaneous norms. In contrast, spontaneous norms within socially distant groups are likely to be marked by a lower level of robustness. In terms of the model, the increase in robustness of norms, i.e., the ability to deal with complexity, must be compensated through an increase in group cohesion, i.e., higher mutual beliefs β_i^k. Highly cohesive social environments have the capacity to sustain robust norms; increasing social distance between actors can be counterbalanced by norms becoming increasingly simpler.

This speculative finding seems to be corroborated in legal anthropology. As previously indicated, methods of social organization akin to what we conceptualize as spontaneous norms have been pervasive in many areas of “primitive” and “archaic” law, a fact certified by many scholars in the field (e.g., Hoebel, 1967; Diamond, 1971; Hallaq, 2004). They develop as customary, unwritten, de-facto rules of behavior, deriving legitimacy from the fact that individuals predominantly comply and have complied in the past.

Significantly, societies exhibiting this kind of social organization also tend to be characterized by low levels of specialization and hierarchy in the administration of social rules,

i.e., promulgation, interpretation, rule change, and dispute adjudication. Conflicts are generally resolved with limited involvement of legal officials (if there exist any), with the involvement of third parties often limited to facultative go-betweens attempting to reconcile conflicted parties. The guiding intellectual principle in these conflict resolution endeavors is the restoration of social harmony and a sense of equity (Evans-Pritchard, 1940; Hoebel, 1967; Merry, 1984). Moreover, the enforcement of these customary rules relies on various networks of alliances and loyalties, facilitating mutual aid. Among fundamental alliances of this nature are descent groups (lineages and clans) and cognatic groups (see, Murray, 1977).[10] Importantly, the abovementioned groups often bear collective responsibility for the transgressions committed by individual members (Parisi & Dari-Mattiaci, 2004; Greif, 2004).[11]

Notably, rules applicable to disputes are contingent on the social standing of the parties involved (Diamond, 1971; Pershi, 1977). The harm inflicted upon members of the same family, clan, or other social group is treated differently from harm directed at outsiders, marking a "distinct opposition of intragroup and intergroup norm" (Pershi, 1977, p. 410). In contemporary discussions surrounding "customary justice" in developing countries, this difference is frequently articulated in the context of norms regulating sexuality, where rights, liberties, and powers held within families and extended families differ from corresponding legal positions held vis-à-vis outsiders (e.g., Institute for Security Studies, 2009).

5. Concluding remarks

The paper has formulated a game-theoretical model of spontaneous norms. Its contribution to the literature lies in considering private assessments among heterogeneous

[10] Descent groups can be defined as groups formed on the basis of descent (demonstrated, assumed, or fictive) from a common ancestor. They are corporate in the sense that they survive the departure of any individual member. In contrast, cognatic groups are personal, i.e., can be defined by the common relationship all its members have with an individual whose death (or other kind of departure) dissolves the group. For a nuanced discussion, see, Murray, 1997.

[11] A crucial aspect to emphasize is the hierarchical nature of clan identities. Clans exhibit internal divisions into subclans and lineages, and can themselves be parts of larger structures such as tribes. Each unit's primary function is to provide aid to members in disputes with non-members: brothers unite in case of a dispute with cousins; cousins and brothers unite in case of a dispute with second cousins, and so forth (see Evans-Pritchard, 1940; Diamond, 1971; Weiner, 2013).

agents – i.e., a scenario wherein individuals maintain diverse views on desirable and undesirable behavior and are not perfectly aware of each other's views. The analysis, situated within the perfectly rational setting, leads to a crucial conclusion: the initial structure of mutual beliefs plays a fundamental role in shaping the long-term equilibrium.

We claim that the model holds relevance across various domains, including general legal theory, the theory of economic institutions, legal anthropology and history, and the sociology of social norms. It also suggests additional research avenues. Most importantly, because the paper presents spontaneous norms as one possible solution to the dual problem of coordination and compliance, and highlights multiple limitations of spontaneous norms, a compelling extension of the analysis would involve comparing such norms with alternative methods of social coordination. As indicated throughout the text, law – interpreted by many theorists as a social convention legitimizing the coordinative efforts of officials – presents an alternative method for dealing with said problem. Hence, a comparison between the two methods naturally emerges.

References

Acemoglu, D., & Jackson, M. O. (2014). History, Expectations, and Leadership in the Evolution of Social Norms. *The Review of Economic Studies*, 82(2), 423-456.

Allen, D., & Barzel, Y. (2011). The Evolution of Criminal Law and Police during the Pre-modern Era. *The Journal of Law, Economics, and Organization*, 27(3), 540-567.

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

Axelrod, R. (1986). An Evolutionary Approach to Norms. *The American Political Science Review*, 80(4), 1095-1111.

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science* 211, 1390-1396.

Bederman, D. J. (2001). *International Law in Antiquity*. Cambridge: Cambridge University Press.

Bertolini, D. (2016). On the Spontaneous Emergence of Private Law. *Canadian Journal of Law & Jurisprudence*, 29(1), 5-36

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.

Bicchieri, C. (2016). *Norms in the Wild. How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.

Bicchieri, C. et al. (2018). "Social Norms", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta and Uri Nodelman (eds.), https://plato.stanford.edu/archives/win2018/entries/social-norms

Bicchieri, C., & Sontuoso, A. (2020). Game-theoretic accounts of social norms: the role of normative expectations. In: Capra, C. M., et al. (eds.), *Handbook of Experimental Game Theory*. Cheltenham: Edward Elgar Publishing, 241-255.

Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3), 337-356.

Carugati, F. et al. (2015). Building Legal Order in Ancient Athens. *Journal of Legal Analysis*, 7(2)2, 291-324.

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

Cooter, R. D. (1996). Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant. *University of Pennsylvania Law Review*, 144(5), 1643-1696.

Cronin, B. (1999). *Community under Anarchy: Transnational Identity and the Evolution of Cooperation*. Columbia University Press.

De Geest, G. (2020). Old Law Is Cheap Law. Washington University in St. Louis Legal Studies Research Paper No. #20-07-05, Available at SSRN: https://ssrn.com/abstract=3704049

Diamond, A. S. (1971). *Primitive Law, Past and Present*. London: Methuen & Co.

Drew, K. F. (1995). Public vs. Private Enforcement of the Law in the Early Middle Ages: Fifth to Twelfth Centuries, *Chicago-Kent Law Review* 70, 1583-1592.

Druzin, B. H. (2016). Social Norms as Substitute for Law. *Albany Law Review*, 79(10), 67-100.

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

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

Evans-Pritchard, E. E. (1940). *The Nuer: A description of the modes of livelihood and political institutions of a Nilotic people*. Oxford: Oxford University Press.

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

Fon, V., & Parisi, F. (2003). Reciprocity-Induced Cooperation. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, 159(1), 76-92.

Fon, V., & Parisi, F. (2007). On the optimal specificity of legal rules. *Journal of Institutional Economics*, 3(2), 147-164.

Ginsburg, T., & McAdams, R. H. (2004). Adjudicating in Anarchy: An Expressive Theory of International Dispute Resolution. *William and Mary Law Review* 45, 1229-1339.

Greif, A. (2004) Impersonal Exchange without Impartial Law: The Community Responsibility System. *Chicago Journal of International Law*, 5(1), 109-138.

Greif, A., & Tabellini, G. (2017). The clan and the corporation: Sustaining cooperation in China and Europe. *Journal of Comparative Economics*, 45(1), 1-35.

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

Hadfield, G. K., & Weingast, B. R. (2012). What is Law? A Coordination Account of the Characteristics of Legal Order. *The Journal of Legal Analysis*, 4(1), 471-514.

Hadfield, G. K., & Weingast, B. R. (2013). Law without the State. Legal Attributes and the Coordination of Decentralized Collective Punishment. *Journal of Law and Courts*, 1(1), 3-34.

Hallaq, W. B. (2004). *The Origins and Evolution of Islamic Law*. Cambridge: Cambridge University Press.

Hart, H. L. A. (1994). *The Concept of Law* (1961). Oxford: Clarendon Press.

Hayek, F. A. (1982). *Law, Legislation, and Liberty. A new statement of the liberal principles of justice and political economy. Volume II: The Mirage of Social Justice* (1976). London: Routledge & Kegan Paul.

Hoebel, E. A. (1967). *The Law of Primitive Man. A Study in Comparative Legal Dynamics*. Cambridge, MA: Harvard University Press.

Institute for Security Studies (2009). Monograph No 159: The Criminal Justice System in Zambia. Enhancing the Delivery of Security in Africa, African Human Security Initiative. Chapter 6: Customary justice.

Kaplow, L. (1992). Rules versus Standards: An Economic Analysis. *Duke Law Journal*, 42(3), 557-629.

Kaplow, L. (1995). A Model of the Optimal Complexity of Legal Rules. *Journal of Law, Economics, & Organization*, 11(1), 150-163.

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

Koyama, M. (2014). The law & economics of private prosecutions in industrial revolution England. *Public Choice*, 159(1/2), 277-298.

Lefkowitz, D. (2017). What makes a social order primitive? In defense of Hart's take on international law. *Legal Theory, 23*(4), 258-282.

Mackie, G. (1996). Ending Footbinding and Infibulation: A Convention Account. *American Sociological Review*, 61(6), 999-1017.

Mahoney, P. G., & Sanchirico, C. W. (2003). Norms, Repeated Games, and the Role of Law. *California Law Review*, 91(5), 1281-1329.

Mahoney, P. G., & Sanchirico, C. W. (2005). General and Specific Legal Rules. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, 161(2), 329-346.

McAdams, R. H. (2000). Focal Point Theory of Expressive Law. *Virginia Law Review* 86, 1649-1730.

McAdams, R. H. (2009). Beyond the prisoners' dilemma: Coordination, game theory, and law. *Southern California Law Review*, 82(2), 209-258.

McAdams, R. H., & Nadler, J. (2005). Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game. *Journal of Empirical Legal Studies*, 2(1), 87-123.

McAdams, R. H., & Rasmusen, E. B. (2007). Norms and the Law. In: Polinsky, M., & Shavell, S. M. (eds.). *The Handbook of Law and Economics*. Elsevier Science. Vol. 2, 1573-1618.

Merry, S. E. (1984). Anthropology and the Study of Alternative Dispute Resolution. *Journal of Legal Education*, 34(2), 277-283.

Morsky, B. & Akçay, B. (2019). Evolution of social norms and correlated equilibria. *Proceedings of the National Academy of Sciences*, 116(18), 8834-8839.

Murray, A. C. (1977). *Germanic Kinship Structure. Studies in Law and Society in Antiquity and the Early Middle Ages*. Toronto: Pontifical Institute of Mediaeval Studies.

Okada, I. (2020). A Review of Theoretical Studies on Indirect Reciprocity. *Games* 11, 27.

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

Parisi, F. (1995). Toward a Theory of Spontaneous Law. *Constitutional Political Economy* 6, 211-231.

Parisi, F. (2000). The Formation of Customary Law. Paper presented at the 96th Annual Conference of the American Political Science Association.

Parisi, F., & Dari-Mattiacci, G. (2004). The rise and fall of communal liability in ancient law. *International Review of Law and Economics*, 24(4), 489-505.

Pershits, A. I. (1977). The Primitive Norm and Its Evolution. *Current Anthropology*, Vol. 18, No. 3 (Sep., 1977), 409-417.

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

Picker, R. C. (1994). An Introduction to Game Theory and the Law. Coase-Sandor Institute for Law & Economics Working Paper No. 22.

Postema, G. J. (1982). Coordination and Convention at the Foundations of Law. *The Journal of Legal Studies*, 11(1), 165-203.

Postema, G. J. (2012). Custom, Normative Practice, and the Law. *Duke Law Journal* 62, 707-738.

Raz, J. (1994). “The Politics of the Rule of Law”. In: *Ethics in the Public Domain. Essays in the Morality of Law and Politics*. Oxford: Oxford University Press, 370-379.

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

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

Sugden, R. (1989). Spontaneous Order. *The Journal of Economic Perspectives*, Vol. 3, No. 4, 85-97.

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

Ullmann-Margalit, E. (1977). *The Emergence of Norms*. New York: Oxford University Press.

Weiner, M. S. (2013). *The Rule of the Clan: What an Ancient Form of Social Organization Reveals About the Future of Individual Freedom*. New York: Farrar, Straus, & Giroux.

Young, P. (2001). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (1998). Princeton, NJ: Princeton University Press.

Young, P. (2015). The Evolution of Social Norms. *Annual Review of Economics*, 7(1), 359-387.