

Evolutionary stable correlation as a core problem of social ontology

Deliberational mechanisms of equilibrium emergence

Valerii Shevchenko

2023

Table of Contents

Evolutionary stable correlation as a core problem of social ontology **1**

Introduction 1

Institutions vs. norms vs. conventions 4

Correlation and asymmetry of strategies 9

The problem with correlation 9

Conclusion 9

References 9

Evolutionary stable correlation as a core problem of social ontology

Introduction

In this paper, I argue that the emergence of evolutionary stable correlation is the core issue of naturalistic social ontology. According to rules-in-equilibria theory, social institutions are the central unit of social ontology (Guala, 2016), and coordination is its main mechanism rooted in evolution (Shevchenko, 2023). As institutions are normatively-driven self-sustaining behavioral regularities designed to solve coordination problems (Aoki, 2007), they share many features with ‘animal conventions’ that help animals solve coordination problems and maintain stable relationships (Hindriks & Guala, 2015). Consequently, understanding the emergence of social institutions requires an examination of the evolutionary mechanisms that enable correlation of strategies with normative force as a key characteristic.

To expand, let us first look at Guala’s (2016) argument that has the following logic:

1. social institutions are backed not by constitutive rules of the form “X counts as Y in (the context of) C”, like in Searle (1995),¹ but

¹For example, “bills issued by the Bureau of Engraving and Printing (X) count

- by regulative rules of the form “do X if Y”
2. from a game-theoretic point of view, regulative rules can be seen as agents’ strategies that comprise a *correlated equilibrium*²
3. constitutive rules are linguistically transformed regulative rules with added theoretical term that represents a certain equilibrium
4. at the same time, many animal species including baboons, lions, swallowtails, and others exhibit behavioral patterns describable in the form of correlated equilibrium, as well (Maynard Smith, 1982)
5. despite the similarity of mathematical representation, the cases of ‘animal conventions’ and human social institutions differ in scope of actionable signals. Building on Sterelny (2003), Guala puts forward an idea that humans can invent and follow new rules, whereas animals are bound to genetically inherited sets of behavioral responses
6. the arbitrariness of rules that humans can invent and follow is grounded in and ontologically depends on shared representations of a given community
7. put differently, the difference in scope of actionable signals between animals and humans can be explained by humans having social epistemology that grounds social ontology.

Although sound, this argument has an Achilles heel: the evolutionary roots of correlation of strategies as the basis of any self-sustaining social coordination, human or not, are still obscure and underdeveloped.

Guala and Hindriks base their account on Maynard Smith’s, who does not use the notion of correlated equilibrium explicitly and discusses what he calls a *bourgeois equilibrium* — a situation in animal territorial behavior, when the most optimal strategy for an animal is to fight for a territory it “owns” it or not fight otherwise. This game is represented in the matrix below.

Guala and Hindriks interpret bourgeois equilibrium as a correlated one. However, there are at least two interpretations of it: *correlated equilibrium* and *evolutionary stable strategy* (ESS) based on uncorrelated

as dollars (Y) in the United States (C)“ (Searle, 1995, p. 28).

²Correlated equilibrium is a general solution concept introduced by Aumann (1974, 1987). As opposed to the classic Nash equilibrium, where players choose their strategies independently, here players choose strategies based on a public signal the value of which they assess privately, thus coordinating their actions according to a given correlation device.

	Hawk	Dove	Bourgeois
Hawk	$\frac{(V-C)}{2}, \frac{(V-C)}{2}$	$V, 0$	$\frac{(V-C)}{2}, V - C$
Dove	$0, V$	$\frac{V}{2}, \frac{V}{2}$	$0, \frac{V}{2}$
Bourgeois	$C - V, \frac{(C-V)}{2}$	$\frac{(C-V)}{2}, C - V$	$\frac{(C-V)}{2}, \frac{(C-V)}{2}$

Table 1: A game-theoretic matrix for a "hawk-dove-bourgeois" game from Maynard Smith's book "Evolution and theory of games". In this game, two players (represented by rows and columns) can choose to be either a hawk (fight for resources), dove (submit and share resources), or bourgeois (submit only when opponent is also bourgeois). The payoffs are determined by the value of the resource (V) and the cost of fighting (C). The table shows the payoff for each player given their own strategy and their opponent's strategy.

asymmetry. They are mathematically distinct, and we will look at both in detail later.

The presented ambiguity creates tension at the backbone of Guala's argument. It means that:

- either 'animal conventions' are mathematically different from human social institutions, for they represent ESS and not correlated equilibrium, and there comes the burden of showing how the former becomes the latter in the course of evolution;
- or that 'animal conventions' are themselves correlated, and there comes the burden of showing how humans acquired the capacity for social epistemology that ontologically grounds social ontology as rules-in-equilibria.

Taking into account the wealth of research on transition from ESS to correlation in game theory (Herrmann & Skyrms, 2021; Kim & Wong, 2017; Lee-Penagos, 2016; Metzger, 2018; Skyrms, 1994), the first option in resolving the tension in Guala's argument becomes insufficient. The transition from ESS to correlation does not intrinsically presuppose the emergence of intentional compliance to norms, as in social institutions, which are normatively-driven and at the same time arbitrary, as will be covered later. Consequently, it will be needed to account for the second option, but to begin, we need to figure out whether social institutions indeed necessitate correlation of strategies. In this paper, I will address the source of the issue—Maynard Smith's notion of bourgeois equilibrium and its interpretations in regard to social coordination.

It is relevant, for if social institutions have emerged from 'animal con-

ventions’ with the aid of cognitive capacities like mindreading and/or mindshaping (Zawidzki, 2013), it constrains social ontology as the scope of possible objects of study to the logical derivatives of social institutions and social coordination in general as discussed in Shevchenko (2023).

This paper is structured as follows. First, it discusses the relationship between social institutions, conventions, and norms, and how conventions emerge through Skyrms’s deliberational dynamics and Harms’s evolutionary functionalism. Second, it examines the correlation and asymmetry of strategies in the emergence of social institutions and explains what correlated equilibrium and uncorrelated asymmetry mean. Two views on correlation versus asymmetry are also discussed. Third, the paper explores the problem with correlation in social institutions as evolved correlated equilibria. It analyzes Guala’s argument about the difference in scope of actionable signals in animals versus humans and Skyrms’s interpretation of Maynard Smith’s “bourgeois” concept. Fourth, it delves into the tension between bourgeois and correlated equilibria with a formal distinction between mixed strategy and correlated equilibria, as well as addressing where randomization of strategies comes from.

Institutions vs. norms vs. conventions

Let us start with the notion of social institutions and move backwards. According to Guala (2016), institutions are rules-in-equilibria, normatively-driven behavioral regularities represented as correlated equilibria. “Rules” here are the recipes guiding and prescribing certain behavior and are *used by the agents themselves*, and “equilibria” are objective stable states of the strategic interaction between agents and population thereof. Other scholars pinpoint normative and self-sustaining nature of institutions. They are “humanly devised constraints that shape human interactions” (North, 1990), “norm-governed social practices” (Tuomela, 2013) and “self-sustaining salient behavioral patterns” (Aoki, 2007). It can be seen that institutions combine “subjective” and “objective” components: they are driven by social norms, that might vary from one population to another, and, at the same time, constrain possible actions and sustain itself.

If social norms are inherently important to institutions, what are they,

and how do they differ from social institutions? According to Bicchieri (2005), social norms are shared expectations, or “rules”, about how people should behave in a given context. These expectations can be either prescriptive, telling individuals what they ought to do, or descriptive, reflecting what most people actually do. Social norms can be modeled as a set of rules or constraints that guide individual behavior. For example, let X be the set of all possible behaviors that an individual can choose from in a given situation. A social norm N can then be represented as a subset of X that specifies which behaviors are considered acceptable or desirable by the group: $N \subseteq X$. The power of social norms lies in their ability to shape behavior without the need for formal enforcement mechanisms like laws or explicit regulations. Individuals often conform to social norms because they want to fit in and be accepted by their peers, or because they believe that following the norm is the right thing to do. Thus, norms are shared expectations about behavior in certain situations and institutions are behavioral patterns that are governed by such shared expectations.

The further required distinction to be made is that of institutions, norms, and conventions. But what are conventions in the first place? Lewis (2008) defines conventions as regularities in behavior that are mutually expected and mutually beneficial for the agents involved. In other words, conventions are shared expectations about behavior that result in cooperative outcomes. To illustrate this concept, Lewis uses the example of driving on the right or left side of the road. This convention is mutually expected because everyone understands that it is necessary for traffic to flow smoothly and avoid accidents. It is also mutually beneficial because if everyone follows the convention, then there is a reduced risk of accidents and delays. Lewis also distinguishes between two types of conventions: coordination conventions and strategic conventions. Coordination conventions are those where agents need to coordinate their actions to achieve a common goal, such as deciding which side of the road to drive on. Strategic conventions are those where agents need to make strategic choices based on what they expect others to do, such as deciding whether to use a turn signal while driving.

For example, consider the following coordination game:

	Drive on left	Drive on right
Drive on left	(1,1)	(-1,-1)
Drive on right	(-1,-1)	(1,1)

In this game, two drivers must choose whether to drive on the left or right side of the road. The payoffs indicate how well each driver does depending on their choice and their partner's choice. If both drivers choose the same side (either both drive on the left or both drive on the right), they each receive a payoff of 1. If they choose different sides (one drives on the left while the other drives on the right), they each receive a payoff of -1 . This game has two pure strategy Nash equilibria: [9] both drivers driving on the left or both driving on the right. In other words, if both drivers follow these conventions, they will achieve a mutually beneficial outcome. Lewis argues that conventions can emerge in situations like this through repeated interactions between agents who learn to coordinate their behavior over time. As more people adopt a particular convention, it becomes more costly for others to deviate from it because they risk being penalized by their partners.

If conventions are mutually expected and mutually beneficial behavioral regularities, how are they different from both social norms and social institutions? O'Connor (2019) draws two crucial distinctions, namely between conventions and social norms, and between more and less arbitrary conventions. The initial distinction implies that not all behavioral regularities possess normative force, meaning that conventions and norms are not that the same. For instance, friends may have a convention of meeting every Friday evening at a bar, and failing to show up does not necessarily imply a violation of a norm. However, when two cars are driving in the same direction towards each other on the same side of the road, the drivers are compelled to swerve to avoid collision. Failing to do so may result in fines or even accidents; hence, swerving becomes an obligatory normative action.

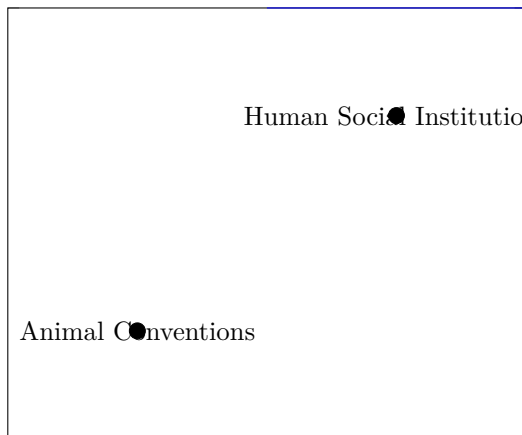
Furthermore, as Bicchieri (2005) asserts, conventions differ from social norms in their association with self-interest and common interest. While they converge with self-interest, they do not necessarily coincide with common interest. In the case of friends gathering at a bar, there is minimal or no tension between self-interest and common interest; however, when driving cars on the road, there is an inherent tension between these interests. O'Connor notes that conventions and norms exist along a continuum, where conventions can acquire normative force based on their position on this spectrum.

The second distinction pertains to the arbitrary and historically contingent nature of conventions, with the recognition that they are subject

to variation and could have been otherwise. This arbitrariness is a fundamental characteristic of conventions, as posited by Lewis. However, Gilbert (1992) has critiqued Lewis's work, noting that not all potential resolutions to a coordination problem offer equal benefits for participants. Hence, where one mode of coordination is more desirable than another, conventionality is not entirely arbitrary. To put it differently, arbitrariness in the context of conventions illustrates a continuum ranging from contingency to necessity. For example, signaling among vervet monkeys may be construed as a convention in the Lewisian sense of recurrent behavioral patterns resolving coordination issues (cf. Harms, 2004; Skyrms, 2010). Nevertheless, this conventionality is not historically contingent insofar as multiple solutions are equally remunerative since adaptive dynamics breaks the symmetry between equilibria. Agents may be genetically predisposed towards certain strategies. Some conventions as more functional and others as more arbitrary.

Putting this into a perspective: - 'animal conventions' are more functional conventions where "normativity", if exists, is grounded in genetically inherited behavioral predispositions; - social institutions are more arbitrary conventions where normativity is grounded in advanced cognitive capacities like mindreading

Genetically Inherited



Necessary

Essentially, social institutions are norm-driven conventions that require certain cognitive capacities which make recognition, complying to and

changing of social norms possible[10]. Three questions arise:

- if institutions are evolved ‘animal conventions’, how do the latter evolve themselves?
- do simple ‘animal conventions’ and social institutions evolve by the same evolutionary mechanism?

Let us first consider two views on evolution of simple ‘animal conventions’: Skyrms’ deliberational dynamics and Harms’ evolutionary functionalism about conventions.

- [(Skyrms, 1994)], [[Vanderschraaf & Skyrms 1993]]
- [(Harms, 2004)]
 - Any semantic convention also referred to as a “rule”, might be considered as a “function-stabilizing mechanism” since it helps to coordinate the behaviour of different organisms or different parts of an organism to perform an evolutionary adapted biological function (Harms, 2004). Fundamentally, rules are maps from conditions to processes — they say what to happen next. Rules for evolutionary adapted traits (AT) might be expressed as follows:
 - $R_{AT} = \{ \langle \text{condition}, \text{process} \mid AT \text{ was selected for performing } \mathbf{process} \text{ in } \mathbf{condition} \} \}$
 - This reads as “the rule for AT is the set of all ordered pairs with a condition in the first place and a process in the second such that AT was selected for performing the process in the conditions” (Harms: 203). Two sub-rules might be derived — for interpretation of a signal and for signal-production. A rule for interpretation for any response mechanism C is expressed as:
 - $R(\text{interpretation})_C = \{ \langle s, b \rangle \mid C \text{ has been selected for } b\text{-ing when } s \text{ is received} \}$ (Harms 2004: 204
 - [[CEU project proposal 2020.pdf]]
- [(Baraghith, 2019)]

As has been shown, conventions are said to be functional. But if social institutions are ‘advanced’ conventions with added cognitive capacities to allow normativity, does this functionality stretch to institutions? If yes, it would mean that conventions and institutions evolve by the same mechanism. And if they do evolve by the same mechanism, the question

is what ensures the emergence of cognitive capacities responsible for normativity?

Hindriks & Guala (2021) claim that institutions have two main functions: etiological and teleological ones, where the first is causal and explains why they persist, and the second is evaluative and explains its purpose. The authors argue that the etiological function of institutions is to promote cooperation and the teleological function is to secure values by means of social norms, as institutions might be seen as norm-governed social practices.

Hindriks and Guala build their account of functions of institutions on the basis of Wright's (1973) analysis of biological functions, which might be summarized in two main conditions. The first is that the function F of an entity A is the cause of the existence of A . The second condition is that F is a consequence of the existence of A , which means that F is non-redundant to the existence of A . The same logic, as the authors argue, applies to institutions. Promotion of cooperation to solve coordination problems is presumably the cause of the persistence of institutions. And securing the normativity of institutions is their purpose.

Correlation and asymmetry of strategies

The problem with correlation

Conclusion

References

- Aoki, M. (2007). Endogenizing institutions and institutional changes*. *Journal of Institutional Economics*, 3(1), 1–31. <https://doi.org/10.1017/S1744137406000531>
- Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1), 67–96. [https://doi.org/10.1016/0304-4068\(74\)90037-8](https://doi.org/10.1016/0304-4068(74)90037-8)
- Aumann, R. J. (1987). Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, 55(1), 1. <https://doi.org/10.2307/1911154>
- Baraghith, K. (2019). Emergence of Public Meaning from a Teleose-

- mantic and Game Theoretical Perspective. *Journal of Philosophy*, 30.
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616037>
- Gilbert, M. (1992). *On Social Facts*. Princeton University Press. <https://books.google.com?id=yYvcDwAAQBAJ>
- Guala, F. (2016). *Understanding institutions: The science and philosophy of living together*. Princeton University Press.
- Harms, W. F. (2004). *Information and Meaning in Evolutionary Processes*. Cambridge University Press. <https://books.google.com?id=zt199e9ugtAC>
- Herrmann, D. A., & Skyrms, B. (2021). Invention and Evolution of Correlated Conventions. *The British Journal for the Philosophy of Science*. <https://doi.org/10.1086/717161>
- Hindriks, F., & Guala, F. (2015). Institutions, rules, and equilibria: A unified theory*. *Journal of Institutional Economics*, 11(3), 459–480. <https://doi.org/10.1017/S1744137414000496>
- Hindriks, F., & Guala, F. (2021). The functions of institutions: Etiology and teleology. *Synthese*, 198(3), 2027–2043. <https://doi.org/10.1007/s11229-019-02188-8>
- Kim, C., & Wong, K.-C. (2017). *Evolutionarily Stable Correlation*. 33(1), 40.
- Lee-Penagos, A. (2016). *Learning to coordinate: Co-evolution and correlated equilibrium* (Working Paper No. 2016-11). CeDEX Discussion Paper Series. <https://www.econstor.eu/handle/10419/163012>
- Lewis, D. (2008). *Convention: A Philosophical Study*. John Wiley & Sons. <https://books.google.com?id=GgCkLtTqBsMC>
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.
- Metzger, L. P. (2018). Evolution and correlated equilibrium. *Journal of Evolutionary Economics*, 28(2), 333–346. <https://doi.org/10.1007/s00191-017-0539-z>
- North, D. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- Searle, J. (1995). *The Construction of Social Reality*. Simon and Schuster. <https://books.google.com?id=zrLQwJCcoOsC>
- Shevchenko, V. (2023). Coordination as Naturalistic Social Ontology: Constraints and Explanation. *Philosophy of the Social Sciences*, 004839312211504. <https://doi.org/10.1177/00483931221150486>

- Skyrms, B. (1994). Darwin Meets the Logic of Decision: Correlation in Evolutionary Game Theory. *Philosophy of Science*, 61(4), 503–528. <https://doi.org/10.1086/289819>
- Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford University Press.
- Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition*. Blackwell.
- Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press. <https://books.google.com?id=6ltpAgAAQBAJ>
- Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. The MIT Press. <https://doi.org/10.7551/mitpress/8441.001.0001>

