

Emergence of evolutionary stable correlation
as a core problem of social ontology
Deliberational mechanisms of equilibrium emergence

Valerii Shevchenko

2023

Table of Contents

Coordination as naturalistic social ontology: constraints and explanation	1
Valerii Shevchenko, HSE University	1
Abstract	1
Keywords: social coordination, social ontology, foundations of social science, naturalism, evolutionary game theory, cognitive mechanisms	2
Introduction	2
Game theory as a model for social ontology	4
Constraints on social ontology: evolution, cognition and physical realization	11
The CNSO framework	13
Equilibrium and causal-mechanistic explanations	17
Conclusion	19
References	19
Introduction	21
1. Social institutions, conventions and norms	22
2. Correlation and asymmetry of strategies	23
3. The problem with correlation	24
4. Conclusion	24

Coordination as naturalistic social ontology: constraints and explanation

Valerii Shevchenko, HSE University

Abstract

In the paper, I propose a project of social coordination as naturalistic social ontology (CNSO) based on the rules-in-equilibria theory of social institutions (Guala & Hindriks, 2015; F. Hindriks & Guala, 2015). It takes coordination as the main ontological unit of the social, a mechanism homological across animals and humans, for both can handle coordination problems: in the forms of ‘animal conventions’ and social institutions, respectively. On this account, institutions are correlated equilibria with normative force. However, if both humans and animals solve coordination problems in a similar way, and only humans have social institutions, how do the latter evolve? I suggest identifying possible causes of this evolution among cognitive capacities like mindreading and imitation by building dynamic models. Social ontology becomes constrained by the evolution of the forms of coordination and by the cognitive mechanisms involved in the emergence and persistence of social institutions. It means that it becomes bound to what might be derived from social institutions, i. e. social roles, structures and their

derivatives. It shows how conceptual relations might emerge from causal. I conclude the paper with the discussion of the relationship between involved types of explanation, mechanistic and equilibrium ones.

Keywords: social coordination, social ontology, foundations of social science, naturalism, evolutionary game theory, cognitive mechanisms

Introduction

Ontology studies “what there is” in the world. Having a viable ontology is vital for the development of any science, for it defines the objects of study. Conjectures about the nature of objects of study influence the ways of approaching them. Philosophy and history of science have seen many cases when mistaken ontology led to scientific mistakes¹. In social science, the issue is even harder due to its massively diversified nature in terms of the main objects of study, used methods, and explanatory strategies².

To get a more objective picture of what there is in the social realm, natural constraints are needed. They might provide evolutionary and cognitive grounds for social phenomena and thus constrain them and narrow down possible objects and methods of study. In other words, natural constraints for social ontology might be a measure to test theories against. There is an array of work in similar directions (Kaidesoja et al., 2019; Turner, 2018; **sperber?**). However, these authors focus on the proximate question—how social phenomena might be related to cognitive ones—instead of an ultimate one—how evolution might have shaped human species to establish unique forms of sociality that are

¹For example, Kuhn (2009, p. 1970) discusses the issue in terms of paradigm change, or change in conceptual networks through the lens of which scientists see the objects of their disciplines. Also, Epstein (2015) discusses ontological mistakes leading to scientific ones in biology and social science (p. 36-49).

²For example, Hawley says that “there are no social analogues of the gravitational waves or Higgs boson which bring physics to the headlines“ (2018, p. 191), and hence, social science cannot generate plausible predictions. Some critics, most notably, Gintis (Gintis, 2007; Gintis & Helbing, 2013), call this diversified state of the social sciences and of sociology in particular “scandalous” and propose to unify them with the evolutionary game theory.

supposed to be a subject of a separate discipline, social science. Let us start with the notion of social ontology and proceed to its natural constraints.

According to Epstein (2018), social ontology divides into two distinct inquiries. The first one deals with the constituents of social entities and addresses the question “what is the social world composed of?”. The second strand of research is concerned with the construction of social categories, or kinds, and with the question “how do social kinds like money, borders, marriage and others get established?”. Individual people constituting a social group exemplify the former inquiry and children playing a game where stuffed animals have a tea party exemplify the latter (Epstein, 2015, p. 57).

The difference between the strands is in the metaphysical relation of individuals and social facts. In the first case, social facts supervene on the facts about individuals, meaning that social properties cannot change without changing the individual ones. In the second case, facts about individuals set up the conditions for something to count as a social fact, i. e. dollar bills with a particular ink and paper and collective acceptance for money. (2015, p. 58).

One of the most famous expressions of the latter is Searle’s (1995) formula “X counts as Y in context C”. For example, “bills issued by the Bureau of Engraving and Printing (X) count as dollars (Y) in the United States (C)” (1995, p. 28). Epstein calls the former relationship grounding and the latter anchoring, where grounding is responsible for instantiation conditions for being, say, money, in a particular context, and anchoring for the mechanism of how these conditions have been established. For example, in Searle’s case, collective acceptance of the fact X is such a mechanism³.

What makes social ontology “naturalistic” in terms of grounds and anchors? I take it to be naturalism about anchors, or a mechanism of bringing about instantiation conditions of a social entity. On this account, social ontology addresses natural constraints of, and their influence on, social phenomena, namely, the role of representations, mental states, their physical realization and evolution.

³The Searle’s formula involves both relations, grounding and anchoring, but for the sake of simplicity and to illuminate collective acceptance as a necessary condition for Searle’s formula I will refer to it as to anchoring.

In the current paper, I provide a framework for a naturalistic social ontology understood as coordination (CNSO) based on Guala's and Hindriks' unified social ontology (Guala & Hindriks, 2015; F. Hindriks & Guala, 2015). Their theory criticizes Searle's view of social ontology by arguing that instantiation conditions for social kinds are brought about by social coordination instead of collective acceptance. Instead of a collective intention to endow a physical entity like a bill the status of money, it is non-intentional historically contingent strategic interaction that results in using something as money because of its utility. Social coordination is understood as correlation of strategies in coordination games expressible as regulative rules of the form "if X, do Y" for each player. If social coordination is an anchor that helps set up what counts as a social fact, the questions of evolutionary constraints on equilibrium emergence and the necessary and sufficient cognitive capacities for it come to the fore. Answering these would provide a naturalistic basis for social ontology.

This paper is structured as follows. The first section views game theory as a modeling tool for social ontology. It reviews a systematic attempt to inform social ontology with game theory proposed by Guala and Hindriks that is called rules-in-equilibrium theory of social institutions (Guala, 2016; F. Hindriks & Guala, 2015). In the second section, I describe the cognitive and evolutionary requirements for emergence and persistence of institutions as the primary entities of social ontology. The implication for social science are discussed, as well. I conclude the paper with the discussion of the form of explanation involved in the proposed framework.

Game theory as a model for social ontology

The tradition of understanding coordination as a source of social order is historically rich. It dates back to Hobbes (2016, p. 1651) with his idea of social order emerging from the rational deliberation of individual agents and resulting into a jointly optimal social decision, an agreement to form a state. The so-called "Hobbesian problem" of social order has been popularized and given the status of the main theoretical problem of sociology by Parsons (1937). The notion of social order presupposes a certain domain ontology given by the relations between agents. As Epstein (2018) points out, a notion of convention was first used as an alternative to agreement by Pufendorf (1673), to refer to

language and law. His point was that conventions do not need to be explicitly agreed to and might exist and work without their intentional design. Later, Hume has famously contributed to the advancement of the notion of convention, which is now often referred to and quoted when coordination problems are involved: “Two men who pull at the oars of a boat, do it by an agreement or convention, tho’ they have never given promises to each other”(Hume, 2003, p. [1740], Bk III, Pt II, Sec II). Hume offered analyses of many social phenomena like money, property, government, justice and others in terms of convention. After Hume, philosophers in the Scottish Enlightenment held that social order is an emergent product of individuals’ interactions, however, no such order has been specifically intended by individuals. As Ferguson (1980, p. 1767) writes, “nations stumble on establishments which are, indeed, the result of human action, but not the execution of any human design”. In other words, tacit conventions as a fundamental part of the human social world have been discussed long before the rise of game theory, and it became especially convenient to treat conventions using game theory.

The study of convention gained particular momentum with Lewis’s seminal book “Convention” (2008, p. [1969]), which, according to Guala (2007) was the first attempt to apply rational choice analysis to the domain of social ontology. Lewis saw conventions as solutions to coordination problems in game-theoretic sense, a basic element of social ontology. But what are conventions and what relation do they bear to coordination?

According to Lewis, social conventions are behavioral patterns emerging from repeated coordination problems between two or more players. The distinctive feature of conventions is that players conform to these patterns, for they expect others to do so, and it is common knowledge that every player is expected to conform. Deviation from a conventional choice of action leads to lower payoff, so players do not have incentives to deviate unilaterally. As conventions are defined in terms of coordination problems, it is useful to elaborate on them. ^1ffa28

Coordination problems are situations where agents have common interests and it is not evident how they can be satisfied. O’Connor (2019) distinguishes two classes of coordination problems, correlative and complementary ones. In correlative coordination problems, agents need to converge on the same choice to coordinate successfully. For example,

consider a driving game, where two players drive towards each other and each can choose the left or right side to drive on. If they both are on the same side and no one swerves, they might crash, and if each of them chooses a different side, they will stay safe. One important feature of this and other coordination problems is arbitrariness, meaning that it does not matter on what side both players would converge. Instead, what matters is that they either coordinate by choosing the same action, for example, swerving to the right. On the game matrix, it is represented as two non-unique equilibria. It means that either of them solves the coordination problem.

Figure 1 about here — driving game

Complementary coordination problems, as opposed to correlative ones, require agents to perform different actions, or strategies, to coordinate successfully. As O'Connor (2019) points out, division of labor and of resources are examples of this class of games. For instance, two roommates want to organize a party and invite guests. To proceed, they need to tidy up the house and order pizza delivery. If they both do the cleaning, there will be no food when the guests come, and if they both order pizza delivery, they will have plenty of food but be embarrassed by the mess at the house.

Figure 2 about here — party game

Regarding conventions, O'Connor (2019) draws two important distinctions: between conventions and social norms, and between more and less arbitrary conventions. First distinction means that not all behavioral regularities have normative force. For example, friends have a convention of meeting each Friday evening at a bar, and showing up is not what each of them strictly ought to do, for if someone does not come, it is fine for the rest of the friends. On the contrary, if two cars are driving on the same side of the road towards each other, the drivers are forced to swerve, for otherwise they might crash. They ought to swerve, for not only might one of them be fined but they might cause an accident. To clarify, as Bicchieri (2005) points out, conventions

are different from social norms in the relationship between self- and common interest. They coincide in the former and do not necessarily coincide in the latter. In the case of friends at a bar, there is no or little tension between self- and common interest, while in the case of driving cars there is. O'Connor stresses that conventions and norms are the poles of a continuum along which the former acquire normative force.
^60075e

The second distinction concerns the arbitrary and historically contingent nature of conventions that they “might have been otherwise”. According to Lewis, this arbitrariness is one of the key distinguishing aspects of conventions. However, as Gilbert (1992) points out in her critique of Lewis’s work, not all possible solutions to a coordination problem are equally profitable for players. In cases where one way of coordinating is more preferred than another, convention will not be that arbitrary. In other words, arbitrariness is a feature of conventions that is a continuum between contingency and necessity. For example, signaling between vervet monkeys might well be modeled as a convention in the Lewisian sense of repeated behavioral patterns of solving coordination problems (cf. Harms, 2004; Skyrms, 2010). However, this convention is not historically contingent in the sense of several possible solutions being equally profitable, for there are evolutionary constraints breaking the symmetry between multiple equilibria. Agents might be hardwired to certain strategies. This distinction, as O'Connor underlines, illuminates some conventions as more functional and others as more conventional⁴.

Another important concept, tightly connected to conventions and norms is that of institution, for it bridges the issue of coordination problems with social science. As famously dubbed by North, social institutions are “the rules of the game in a society or, more formally, the humanly devised constraints that shape human interactions” (North, 1990, pp. 3–4). They are also self-sustaining salient behavioral patterns (Aoki, 2007) and norm-governed social practices (Tuomela, 2013). However, it is not clear how institutions are related to conventions and norms.

Guala (2016) offers a definition of institutions as sets of correlated

⁴Harms (2004) goes even further and defines conventions as a “function-stabilizing mechanism”, meaning that, in evolutionary terms, coordination in any signaling system is meant to promote a function (Harms, 2004, p. 202).

equilibria with normative force. With Hindriks, they propose a “unified social ontology” that views social institutions as rules in equilibria represented by theoretical terms like “money” or “marriage” (F. Hindriks & Guala, 2015). “Rules” are the recipes that guide and prescribe certain behavior and that are used by the agents themselves, and “equilibria” are objective stable states of the strategic interaction between agents and population thereof. Rules-in-equilibria theory bridges the accounts of regulative rules, equilibria of strategic games and constitutive rules of the form: “X counts as Y in C”. The former two are complementary ones, and the latter one supplements it by providing a symbolic representation, or term like “money”. ⁵

The rule-based account conceives of social institutions as rules guiding and constraining behavior in social interaction. In sociology, the tradition of treating institutions as rules dates back to classical figures like Weber (1924) and Parsons (1951, p. 1935). The equilibrium-based account sees institutions as behavioral regularities and, most importantly, solutions to coordination problems. The constitutive rules account, introduced by Searle (1995, 2010), sees institutions as systems assigning statuses and functions to physical entities.

According to the Guala and Hindriks, the rule-based account is insufficient, for it cannot explain why some rules are followed and others not. To address this issue, an equilibrium account is needed to show the strategic character of rule-following. They illustrate this point by comparing the two paradigmatic games from game theory: the prisoner’s dilemma and stag hunt.

Figure 3 about here — PD and stag hunt

Although mutual defection in the prisoner’s dilemma is a Nash equilibrium⁵, it is not a social institution, for it is not self-sustaining due to the independence of players’ strategies. In contrast, the mutual decision to hunt a stag instead of a hare, which are also both Nash equilibria, is actually an institution, for it requires a correlation of players’ strategies to achieve a bigger joint payoff. The latter means that the strategy is

⁵Nash equilibrium is a solution concept describing a strategy profile consisting of each player’s best response to the other player’s strategies where no one gains bigger payoff by deviating unilaterally.

salient and beneficial for players. This, according to Guala and Hindriks, explains why some rules are followed and others not.

However, the notion of players' correlated strategies, or correlated equilibrium⁶, as an *explanans* of the stability of institutions is insufficient, as the authors point out, for it is too loose. They provide an example of non-human animals solving coordination problems, but still not having institutions. For instance, male baboons, lions, swallowtails and other species exhibit a recurring behavioral pattern that can be described in terms of correlated equilibrium, when males patrol an area to mate with females and have ritual fights with intruders if encountered. The evolved pair of players' strategies—"fight if own" and "surrender if do not own"—minimizes possible damage to both parties and lets the owner occupy territory and mate. This problem was originally represented with the "Hawk-dove game" by Maynard Smith (1982). This game is symmetric, meaning that each player has the same action profiles. This is why there is only one number in each cell.

Figure 4 about here — hawk-dove from maynard smith

Maynard Smith calls this solution "bourgeois equilibrium", for it describes animal territorial behavior: "Hawk if owner" and "Dove if intruder".

Guala and Hindriks argue that human and animal conventions differ only in the scope of actionable signals. The set of signals to which animals might respond is rather narrower than that of humans, due to the tight coupling of stimulus and response to achieve coordination. In other words, animals are hardwired to their strategic choices in coordination problems, and human are not. For example, the "hawk" and "dove" strategies in the case of territory ownership in animals is genetically inherited and not arbitrary. It is functional rather than conventional in O'Connor's terms. However, as Sterelny (2003) argues, more complex creatures like humans are able to decouple stimuli and behaviour with the aid of representation of the environment that con-

⁶Correlated equilibrium is a more general solution concept than Nash equilibrium introduced by Aumann (1974, 1987). Players choose their strategies based on some public signal the value of which they assess privately, thus coordinating their actions according to a given correlation device.

ditions behavior. In other words, humans are able to invent and follow different rules given the same correlation device—a source of signal that correlates agents’ actions. Moreover, rules are themselves symbolic representations of strategies of a given game. These representations not only serve as symbolic markers of the properties of equilibria, but considerably save cognitive effort by functioning as a shortcut. “Stop if red” is such a rule that guides behavior and at the same time serves a symbolic representation of an equilibrium. But this alone does not say much about social ontology.

To address this issue, Guala and Hindriks, drawing on F. A. Hindriks (2005), propose to bridge their rules-in-equilibria account of institutions with the constitutive rules account. The latter presents institutions as systems of statuses and functions paradigmatically proposed by Searle (1995) as the formula “X counts as Y in C”. Searle draws a sharp distinction between constitutive and regulative rules, emphasizing the difference in their syntax, for that of the latter is “if X, do Y”. As the authors note, constitutive rules are linguistically transformed regulative rules, aided with a new term to name an institution. Combining these accounts enables researchers to investigate Y-terms like “money” or “marriage” used by individuals in everyday life and analyze their internal regulative and strategic character, thus bridging explicit ontology of social science and implicit ontology of ordinary language.

The unified rules-in-equilibria account has several shortcomings. Although it mentions that social coordination as correlation of strategies with normative forces seems to have evolutionary roots, this approach does not address the constraints shaping animal conventions into social institutions and the role of cognitive mechanisms in it. However, Guala (2020) is concerned with a similar issue and asks, “what cognitive mechanisms establish coordination?”. In addition, Bicchieri (2005; 2018) refers to social norms as being “activated” by scripts and grounded in cognitive schemata. However, its inner workings remain unexplained.

The main issue with game-theoretic explanations of social coordination in terms of equilibria, as Kaidesoja et al. (2019) argue, is that they do not explicate causal processes or mechanisms. Hence, the question to be answered is “what cognitive mechanisms are responsible for the evolutionary transition from animal conventions to human social institutions?” This would require deeper entwinement of game-theory, cognitive science and evolutionary theory.

Constraints on social ontology: evolution, cognition and physical realization

S. Turner (2007) argues that social theory relies on, and makes generous use of, cognitive concepts like “meaning”, “belief” and “collective representations”. However, it uses these concepts only metaphorically and does not engage with the actual cognitive science.

Turner proposes what he calls a “sane” constraint: social theory and social ontology as the focal set of entities and processes, ought to be physically, computationally and cognitively realistic. It means that social concepts and explanations must be grounded in physically realizable cognitive entities and be computationally tractable. If game-theoretic explanations of coordination are computationally tractable, for they provide an insight into the structure of coordination, the other two requirements remain unsatisfied.

Turner (2018), drawing on Guala (2016), sketchily connects the notions of joint attention, mirror neurons, focal points, correlated equilibria and social coordination. He suggests that “<...> we can think of actual societies as made up of multiple focal points which are the subject of joint attention by different overlapping groups, as the distributed rather than centralized source of multiple modes of coordination” (2018, p. 209). Thus, he makes preliminary steps towards meeting his own constraints.

These steps explain coordination describable in game-theoretic terms as grounded in cognitive and neuronal phenomena. They do not comprise a complete explanation yet, but point to a certain direction. In other words, coordination is the main mechanism of scaling up from autonomous agents to larger social groups, which is, itself, grounded in cognitive and neural phenomena like joint attention and mirroring. This idea might be tested with the dynamic models, but first it needs to be narrowed down with the two sets of questions.

First, if normative force is what distinguishes coordination problems in animals and humans, how does it emerge, and what are the minimal cognitive requirements for it? In other words, what human cognitive capacities made possible the emergence of social institutions as rules-in-equilibria? These questions presuppose that cognitive phenomena might have been contributing to the emergence and persistence of social institutions as distinct from animal conventions. This contribution constrains sociological theorizing in a way that ontological units of the social should be bound to the mechanism of coordination as cognitively implemented and physically realized correlation of strategies: to be either the products of coordination or their derivatives. ⁷

The second set of questions is more general and concerns the relationship between the social and cognitive phenomena. If cognitive phenomena are admitted to have explanatory and ontological significance for sociological explanations, two positions are possible. First, it is non-reductive physicalism like Sperber's (2011) that allows ontological reduction without theoretical one. It means that cognitive phenomena have causal powers and real existence, and social ones do not, while remaining useful fictions. On this account, social institutions as rules-in-equilibria might be ontologically reduced to their representations, which have causal powers. However, it also retains explanatory autonomy of the game-theoretic notions of coordination. The second option is reductive physicalism like Turner's (2018; 2019) that presupposes eliminativism about social scientific concepts. It means that the foundational social scientific concepts like "collective representation", "social structure" and "belief" must be abandoned in favor of cognitive and neurophysiological ones, on which the former ones supposedly depend. On this account, naturalistic social ontology is not "social", but consists of neurophysiological states of agents that explain social phenomena⁷.

⁷For example, S. Turner (2019) proposes to supplant Weber's explanation of social action as based on empathic understanding, or *Verstehen*, with a neuroscientific

However, if cognitive mechanisms happen to be involved in the transition from animal conventions to human social institutions, it does not entail ontological reducibility of social institutions to cognitive and neuronal entities like Sperber and Turner propose. Instead, their involvement only assumes their strong influence, not complete determination. Hence, it might be only a proof-of-concept of the close relation of cognitive capacities to social ontology, but it cannot specify the type of the relation and corroborate either non-reductive or reductive physicalism about social entities.

The CNSO framework

To start unfolding the “Coordination as Naturalistic Social Ontology” framework, it is important to note that the notion of coordination is relevant in two meanings, game-theoretic and cognitive scientific ones. In the former case, coordination (C1) is a successful result of agents’ strategies leading to a solution of a game, either optimal or suboptimal. In the latter case, coordination (C2) is also a successful result of a social interaction, either spatiotemporally synchronous or not, that is possible due to the cognitive capacities to represent goals and actions, to monitor and predict those and to adjust one’s own actions to the actions of another individual with whom coordination takes place (Vesper et al., 2010).

Naturally constrained social ontology might be understood as coordination—both a continuing process and its results recursively interacting with each other: small groups, social institutions and social structures⁸. Coordination itself can be understood as a causal mechanism of correlating agents’ strategies by their monitoring, predicting and adjusting to other agent’s actions. Coordination has the global scope of evolutionary time which is computationally tractable

notion of pattern completion inference (Barsalou, 2013 Nov-Dec), meaning that parallel neuronal processing in the brain “complete the patterns” given a social setting. When your friend waves you from the opposite side of the street, a handful of brain modules—memory, face recognition, sensorimotor control and others—process environmental information in parallel. As Turner puts forward, this explains social action, and similar explanations might be given for other social phenomena.

⁸In the preliminary understanding, a process is more ontologically primary than entities like institutions, which means that social *entities* are byproducts of the recursive process of coordination. This idea is closely connected with the status of objects in structural realism (French, 2010), and is out of the scope of the paper.

via dynamic models, and the scope of local social interaction that is empirically testable with the cognitive scientific experiments in joint action research (Knoblich et al., 2011; Sebanz & Knoblich, 2021; Török et al., 2019). This description implies two natural constraints on social ontology as coordination.

The first natural constraint is evolutionary conditions for equilibrium emergence for distinctively human forms of social coordination, which is social institutions as described in rules-in-equilibria account earlier. Namely, if animals are able to solve coordination problems with ‘animal conventions’, which might be described formally as evolutionary stable strategies in coordination and cooperation games, what enables the emergence of social institutions as conventions with normative force? Evolutionary conditions for the emergence of uniquely human coordination constraint social ontology by naturalizing the problem of social order—that social institutions were not created *ex machina*, but emerged gradually from the other forms of coordination that had no full-fledged normative component⁹.

The second, cognitive constraint is closely related to the first one. If social institutions evolved from animal forms of coordination, and humans are drastically distinct from animals in terms of cognitive capacities, these capacities must have influenced the emergence of social institutions. Although this sounds truistic, it nevertheless binds social ontology to these cognitive capacities, meaning that social institutions are ontologically dependent on them. As dependence might be causal or constitutive, and cognitive capacities might have had causal influence on the emergence of social institutions, it follows that institutions might causally depend on cognitive mechanisms¹⁰. Although it does not make social institutions a causal phenomenon *per se*, it says that they are only possible due to cognitive capacities of agents.

To further untangle the relationship between the cognitive and the social and to divide the labor, Sarkia et al. (2020) propose a framework based on mechanistic philosophy of science (Craver & Darden, 2013;

⁹Harms (2004) discusses what he calls “primitive content” in animal signaling, that is signals that simultaneously track the state of the environment and motivate. For example, danger calls of vervet monkeys at the same time convey information that there is an eagle and that it is better to hide. Huttegger (2007) provides a helpful distinction of indicative and imperative signals.

¹⁰See Guala (2016), ch. 11 on this relation. He discusses social institutions as dependent on the human capacity for mindreading.

Glennan, 2017). Its main idea is that mechanisms for phenomena consist of entities “whose activities and interactions are organized so as to be responsible for the phenomenon” (Glennan, 2017, p. 17). Drawing on Bechtel’s (2009) exposition of mechanistic explanation in cognitive science, the authors suggest that such explanations might be given by answering four questions: (1) what is the phenomenon? (2) what entities and processes does it consist of? (3) what are the interactions of these entities contributing to the phenomenon? (4) what is the environment where the mechanism is situated, and how does it affect its functioning? The division of labor, the authors argue, is accomplished by answering different sets of Bechtel’s questions. Evolutionary anthropologists and developmental psychologists answer the first and the fourth of them by studying the key differences between human and great ape cognitive capacities and the development of uniquely human ones in the course of human ontogeny. Cognitive scientists answer the second and third questions by identifying particular cognitive mechanisms, i.e. entrainment and common object affordances (Knoblich et al., 2011). In addition, the authors put forward the idea that social scientists might address further questions outlying those suggested by Bechtel, i.e. complex social networks emergent from social coordination.

CNSO holds that coordination is the main ontological unit of the social—it is an ongoing process governed by a causal mechanism that leads (intentionally or not) to social institutions—normatively-driven self-sustaining behavioral regularities. Social institutions provide familiar sociological terms like roles, and the latter depend on the former constitutively, not causally: if there were no social institution of ‘marriage’, there were no roles of ‘wife’ and ‘husband’ that imply certain rights and duties. These are the first-order derivatives of CNSO, and the notion of social structures as patterns of relationships between agents is its second-order derivative.

This derivation is a preliminary way to address Kincaid’s discussion of social structures as representing conceptual and not causal relations (Kincaid, 2008). If coordination is a causal mechanism, and its successful recurring result is a social institution as rules-in-equilibria, it means that social institutions are an inherently causal phenomenon with an appended theoretical term like “marriage”. This term generates conceptual relations between roles. This would show how causal relations regarding the social establish conceptual, or constitutive ones. Overall, naturalism about what Epstein (2015) calls anchors—instantiation con-

ditions for what to count as a social fact—means that the main ontological unit of the social—coordination resulting into institutions—comes into being causally, and not constitutively. However, these implications must be addressed in a separate paper. In the next section I address a question of relating two different types of explanation involved in the CNSO framework: equilibrium and causal-mechanistic. ^2a88ff

Equilibrium and causal-mechanistic explanations

Depending on its meaning, either C1 or C2, coordination is explained differently. C1, which is game-theoretic notion of correlation of strategies, provides what is called an equilibrium explanation. C2, or cognitive-scientific notion of monitoring, prediction and adjustment to the actions of an opponent, provides a causal-mechanistic explanation, for it provides an “entities and processes” picture. According to Sober, these are in tension, for the former shows how an event might have occurred regardless the actual causes involved, whereas the latter shows how an event was actually produced (1983). Mechanistic and equilibrium explanations differ in the source of their explanatory force: the former show entities and their relations as responsible for a phenomenon, whereas the latter show “deep mathematical structure” and global stability of a phenomenon.

In the current debates, however, equilibrium explanations are seen as a subgroup of either structural, or “distinctively mathematical” explanations or of causal-mechanistic ones (Huneman, 2018; Sperry-Taylor, 2021; Suárez & Deulofeu, 2019). The relation of these two types of explanation is problematic in a semantic sense, for it is not evident what precisely the notion of a mechanism means regarding naturalistic explanation of coordination as social ontology. Is coordination itself a mechanism?

In C2, cognitive capacities for coordination—monitoring, prediction

and adjustment—are explained by fine-grained cognitive mechanisms in a strong sense of entities and their relations as responsible for the occurrence of the phenomenon, for example, entrainment, common object affordances or perception-action matching among others [Knoblich et al. (2011)]¹¹. C2 might be said to be a nested mechanism, for it involves distinct capacities which are mechanisms themselves on a lower level and their relation as responsible for coordination. On the contrary, C1 cannot be said to be a mechanism in the strong sense, for it does not consist of any entities and processes. However, one might object that there are agents and their strategies, but C1's explanation does not gain its force solely from the relation of agents and their strategies in a mechanistic sense. Instead, coordination as an equilibrium outcome comes from accounting for initial conditions and specifying dynamic processes in the studied system. These represent a mathematical rather than empirical structure of coordination. The problem, then, is relating mathematical and empirical structures of the presumably same phenomenon. However, C1's explanation can still be causal without being mechanistic in the strong sense.

Sperry-Taylor (2021) points out, that equilibrium explanations are not monolithic and that they do identify causes. It means that they explain not only by appealing to system's structural relationships but by taking the system's initial conditions and dynamic processes as possible causes. These causes are understood as interventions, or variables subject to manipulation and control on which an outcome depends (Woodward, 2005). Discussing the emergence of social norms, Sperry-Taylor suggests that the introduction of multiple competing equilibria affects possible outcomes and manipulating initial conditions and dynamic processes might lead to emergence of different equilibria. Explanatory power of such explanation comes, as the author notes, from more information that allows to address both selection of a certain equilibrium and its dynamics to disequilibrium and back, whereas canonical equilibrium explanations explain only system's persistence and have nothing to intervene with.

¹¹Entrainment is social motor coordination process, which is temporally synchronous. For example, people applauding in a theatre. Common object affordance is action opportunity of an object. If agents have similar action repertoire, they might engage in spontaneously emerged coordination regarding a certain object. Perception-action matching is a process of matching observed actions and agent's own action repertoire (Knoblich et al., 2011).

Conclusion

Inquiring into the cognitive capacities that might have affected the emergence of equilibria with normative forces presupposes manipulating initial conditions and dynamic processes of a system that lead to emergence of equilibria, and hence it can provide a causal story. The next step would be to “zoom in” and explicate the causally efficacious cognitive capacities, for example, mindreading, in experimental settings. This would provide a naturalistic social ontology understood as coordination and show that social institutions are produced causally, and their derivatives like social roles and social structures are produced conceptually, or logically. This restricts social ontology only to the entities that might be logically derived from social institutions and its immediate derivatives like roles. To vindicate this intuition, it is needed to build dynamic models of the transition from ‘animal conventions’ to social institutions with the help of the major cognitive capacities that are usually said to be uniquely human, for example, mindreading and imitation. Formally, it means relating the game-theoretic concepts of evolutionary stable strategy and correlated equilibrium through the co-evolution of within- and between-organism coordination represented by Lewis signaling game: the more effective inner coordination became, the more efficient the computation and information processing became, and more states of the world were able to be represented, and this, supposedly, influenced the transition from ‘animal conventions’ to social institutions.

References

Introduction

Social institutions are the central unit of social ontology (Guala, 2016) and coordination is its main mechanism that is rooted in evolution (shevchenko2022?). As institutions are normative conventions that solve coordination problems #insertSource , they are structurally similar to ‘animal conventions’ that help animals solve coordination problems (F. Hindriks & Guala, 2015). This means that the main problem with the explanation of the nature of institutions is its emergence as a certain kind of game-theoretic equilibrium from another one. Thus, the question is, what evolutionary and cognitive mechanisms enable correlation of strategies with normative force as the main characteristic of human social institutions. [[Mechanism of attaining of correlated equilibrium in humans might increase explanatory depth of theory of institutions]] [[poss_article#^fbc225]]

However, [[It is hard to figure out a true dynamic trajectory by which coordination problems and socialization co-evolved]] [(Ross, 2008)]

But why this should matter? If social institutions have emerged from ‘animal conventions’ with the aid of cognitive capacities like mindreading or mindshaping #insertSource (zawidski), it constrains social ontology as the scope of possible objects of study to logical derivatives of social institutions as the only truly causal social phenomenon. [[poss_article#^2a88ff]]

This paper is structured as follows: ...

1. Social institutions, conventions and norms

How social institutions as a distinct kind of equilibrium *might* have emerged from ‘animal conventions’ is a twofold question. First, it is essentially philosophical, for it addresses the ultimate cause of the existence of institutions as a specifically human thing. And this pertains to the discussion of functionalism about institutions. Second, it is game-theoretic, for it involves the notion of equilibria, conventions and norms.

How institutions, conventions and norms are related? [\[\[pos Article#1ffa28\]\]](#) [\[\[pos Article#3ac825\]\]](#) [\[\[pos Article#60075e\]\]](#) Why institutions exist? [\[\[Institutions are there for they incentivise which is due to gene-culture coevolution and not only natural selection\]\]](#) [\[\[Institutions have etiological and teleological functions\]\]](#)

This leads to an important question: if social institutions are normatively-driven conventions, how do conventions evolve, even before they are normatively-driven? Two answers, although similar, are possible.

First, it is Skyrms’s deliberational dynamics. [\[\[Vanderschraaf & Skyrms 1993\]\]](#)

Second, it is Harms’s evolutionary functionalism about conventions.

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]]

Adopting this view of conventions to human social institutions is problematic, for it is still not evident that the model of animal signaling can be applied to humans. To show, how institutions might have evolved, it is needed to compare possibility conditions of ‘animal conventions’ and human social institutions.

2. Correlation and asymmetry of strategies

Guala and Hindriks appeal to Maynard Smith (1982) in their exposition of the theory of social institutions as correlated equilibria. They see “Hawk-dove-bourgeois” game of animal territorial ownership as a prototypical ‘animal convention’. According to the authors, the difference, however, is that humans are able to decouple response from stimuli with the representation of environment.

Hence, if both animals and humans solve coordination problems, let us start with the game-theoretic analysis of ‘animal conventions’ and Maynard Smith’s notion of animal coordination as a “Hawk-dove game”.

[(Maynard Smith, 1982)] 1. what is hawk-dove game 2. how bourgeois appears 3. what it means

There are two interpretations of the “Hawk-dove-bourgeois” game. One of them involves correlation, and another does not. But first, it is needed to clarify, what is meant by “correlation of strategies”.

1. What is correlation and why it explains (example) - Guala — social institutions are correlated equilibria - Gintis — social norms are correlated equilibria - Vanderschraaf — conventions as correlated equilibrium - Skyrms — correlation evolves naturally **2. Definition of correlated equilibrium**

In their theory, Guala and Hindriks appeal to Skyrms’s interpretation of the “Hawk-dove” that is with correlation. According to Skyrms, the implicative nature of the “bourgeois” strategy in the form “if own, then Hawk” and “if not own, then Dove” is genuinely correlative. [(Skyrms, 2014)]

However, another interpretation involves uncorrelated asymmetry [(O'Connor, 2019)]. She uses social categories as solutions to inherently institutional coordination problems **3. Definition of uncorrelated asymmetry (+ example)** - O'Connor and Saunders's comment — [(Saunders, 2022)]

The problem is how social institutions as correlation might have evolved from 'animal conventions' if the latter involve or do not involve correlation. And the [[Mechanism of attaining of correlated equilibrium in humans might increase explanatory depth of theory of institutions]]

3. The problem with correlation

Two options for the emergence of social institutions as correlated equilibria with normative forces are available: either from uncorrelated or correlated "Bourgeois" equilibrium.

If "Bourgeois" equilibrium is uncorrelated, then it is a problem of emergence of evolutionary stable correlation — [(Kim & Wong, 2017)], or how correlated equilibrium emerges from ESS.

If "Bourgeois" strategy is correlated, then another question becomes important, namely what Guala and Hindriks cite as difference between humans social institutions and 'animal conventions' as difference in the scope of actionable signals. [[Human and animal conventions differ in scope of actionable signals]]

However, to admit that 'animal conventions' are correlated equilibria as well, it is first needed to answer the question, where does randomization of strategies come from? Or what breaks the symmetry in agents' strategies given that ESSes are symmetric? — [(Metzger, 2018)]

4. Conclusion

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>
- Barsalou, L. W. (2013 Nov-Dec). Mirroring as Pattern Completion Inferences within Situated Conceptualizations. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(10), 2951–2953. <https://doi.org/10.1016/j.cortex.2013.06.010>
- Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, 22(5), 543–564. <https://doi.org/10.1080/09515080903238948>
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616037>
- Bicchieri, C., Muldoon, R., & Sontuoso, A. (2018). Social Norms. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2018/entries/social-norms/>
- Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press. <https://books.google.com?id=ESI3AAAAQBAJ>
- Epstein, B. (2015). *The ant trap: Rebuilding the foundations of the social sciences*. Oxford University Press.
- Epstein, B. (2018). Social Ontology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/summer2018/entries/social-ontology/>
- Ferguson, A. (1980). *An Essay on the History of Civil Society, 1767*. Transaction Publishers.
- French, S. (2010). The interdependence of structure, objects and dependence. *Synthese*, 175(S1), 89–109. <https://doi.org/10.1007/s11229-010-9734-2>
- Gilbert, M. (1992). *On Social Facts*. Princeton University Press. <https://books.google.com?id=yYvcDwAAQBAJ>
- Gintis, H. (2007). A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences*, 30(1), 1–16. <https://doi.org/10.1017/S0140525X07000581>
- Gintis, H., & Helbing, D. (2013). Homo Socialis: An Analytical Core for Sociological Theory. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2362262>
- Glennan, S. (2017). *The New Mechanical Philosophy* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oso/9780198779711.001.0001>

0001

- Guala, F. (2007). The Philosophy of Social Science: Metaphysical and Empirical. *Philosophy Compass*, 2(6), 954–980. <https://doi.org/10.1111/j.1747-9991.2007.00095.x>
- Guala, F. (2016). *Understanding institutions: The science and philosophy of living together*. Princeton University Press.
- Guala, F. (2020). Solving the Hi-lo Paradox: Equilibria, Beliefs, and Coordination. In A. Fiebich (Ed.), *Minimal Cooperation and Shared Agency* (Vol. 11, pp. 149–168). Springer International Publishing. https://doi.org/10.1007/978-3-030-29783-1_9
- Guala, F., & Hindriks, F. (2015). A UNIFIED SOCIAL ONTOLOGY. *The Philosophical Quarterly*, 65(259), 177–201. <https://doi.org/10.1093/pq/pqu072>
- Harms, W. F. (2004). *Information and Meaning in Evolutionary Processes*. Cambridge University Press. <https://books.google.com?id=zt199e9ugtAC>
- Hawley, K. (2018). Social Science as a Guide to Social Metaphysics? *Journal for General Philosophy of Science*, 49(2), 187–198. <https://doi.org/10.1007/s10838-017-9389-5>
- Hindriks, F. A. (2005). *Rules & institutions: Essays on meaning, speech acts and social ontology: Essays over betekenis, taalhandelingen en sociale ontologie = Regels & instituties*. Haveka BV.
- 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>
- Hobbes, T. (2016). *Thomas Hobbes: Leviathan (Longman Library of Primary Sources in Philosophy)* (M. Missner, Ed.; 0th ed.). Routledge. <https://doi.org/10.4324/9781315507613>
- Hume, D. (2003). *A Treatise of Human Nature*. Courier Corporation. https://books.google.com?id=zHYO1Fh9_JMC
- Huneman, P. (2018). Outlines of a theory of structural explanations. *Philosophical Studies*, 175(3), 665–702. <https://doi.org/10.1007/s11098-017-0887-4>
- Huttegger, S. M. (2007). Evolutionary Explanations of Indicatives and Imperatives. *Erkenntnis*, 66(3), 409–436. <https://doi.org/10.1007/s10670-006-9022-1>
- Kaidesoja, T., Sarkia, M., & Hyryläinen, M. (2019). Arguments for the cognitive social sciences. *Journal for the Theory of Social Behaviour*, 49(4), 480–498. <https://doi.org/10.1111/jtsb.12226>
- Kim, C., & Wong, K.-C. (2017). *Evolutionarily Stable Correlation*.

33(1), 40.

- Kincaid, H. (2008). Structural Realism and the Social Sciences. *Philosophy of Science*, 75(5), 720–731. <https://doi.org/10.1086/594517>
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation* (Vol. 54, pp. 59–101). Elsevier. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- Kuhn, T. S. (2009). *The structure of scientific revolutions* (3. ed., [Nachdr.]). Univ. of Chicago Press.
- 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.
- O'Connor, C. (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution* (First edition). Oxford University Press.
- Parsons, T. (1937). *The Structure of Social Action*. The Free Press.
- Parsons, T. (2015). The Place of Ultimate Values in Sociological Theory. *The International Journal of Ethics*. <https://doi.org/10.1086/intejethi.45.3.2378271>
- Pufendorf, S. (1673). *De Officio Hominis et civis juxta legem naturalem libri duo*. Lund: Junghans. https://catalogue-bu.u-bourgogne.fr/discovery/fulldisplay/alma991001697579706659/33UB_INST:33UB_INST
- Ross, D. (2008). Classical Game Theory, Socialization and the Rationalization of Conventions. *Topoi*, 27(1-2), 57–72. <https://doi.org/10.1007/s11245-008-9028-1>
- Sarkia, M., Kaidesoja, T., & Hyyryläinen, M. (2020). Mechanistic explanations in the cognitive social sciences: Lessons from three case studies. *Social Science Information*, 59(4), 580–603. <https://doi.org/10.1177/0539018420968742>
- Saunders, D. (2022). How to Put the Cart Behind the Horse in the Cultural Evolution of Gender. *Philosophy of the Social Sciences*, 52(1-2), 81–102. <https://doi.org/10.1177/00483931211049770>
- Searle, J. (1995). *The Construction of Social Reality*. Simon and Schuster. <https://books.google.com?id=zrLQwJCcoOsC>

- Searle, J. (2010). *Making the Social World: The Structure of Human Civilization*. Oxford University Press. <https://books.google.com?i d=kz6R0eDZ5OEC>
- Sebanz, N., & Knoblich, G. (2021). Progress in Joint-Action Research. *Current Directions in Psychological Science*, 30(2), 138–143. <https://doi.org/10.1177/0963721420984425>
- Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford University Press.
- Skyrms, B. (2014). *Evolution of the social contract* (Second edition). Cambridge University Press.
- Smith, J. M. (1982). *Evolution and the Theory of Games*. Cambridge University Press. <https://books.google.com?id=Nag2IhmPS3gC>
- Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43(2), 201–210. <https://doi.org/10.1007/BF00372383>
- Sperber, D. (2011). A naturalistic ontology for mechanistic explanations in the social sciences. In P. Demeulenaere (Ed.), *Analytical Sociology and Social Mechanisms* (pp. 64–77). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921315.004>
- Sperry-Taylor, A. T. (2021). Reassessing equilibrium explanations: When are they causal explanations? *Synthese*, 198(6), 5577–5598. <https://doi.org/10.1007/s11229-019-02423-2>
- Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition*. Blackwell.
- Suárez, J., & Deulofeu, R. (2019). Equilibrium Explanation as Structural Non-Mechanistic Explanations: The Case of Long-Term Bacterial Persistence in Human Hosts. *Teorema: Revista Internacional de Filosofía*, 38(3), 95–120. <https://www.jstor.org/stable/26874514>
- Török, G., Pomiechowska, B., Csibra, G., & Sebanz, N. (2019). Rationality in Joint Action: Maximizing Coefficiency in Coordination. *Psychological Science*, 30(6), 930–941. <https://doi.org/10.1177/0956797619842550>
- Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press. <https://books.google.com?id=6ltpAgAAQBAJ>
- Turner. (2018). *Cognitive Science and the Social: A Primer* (1st ed.). Routledge. <https://doi.org/10.4324/9781351180528>
- Turner, S. (2007). Social Theory as a Cognitive Neuroscience. *European Journal of Social Theory*, 10(3), 357–374. <https://doi.org/10.1177/1368431007080700>
- Turner, S. (2019). Verstehen Naturalized. *Philosophy of the Social*

Sciences, 49(4), 243–264. <https://doi.org/10.1177/0048393119847102>

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8-9), 998–1003. <https://doi.org/10.1016/j.neunet.2010.06.002>

Weber, M. (1924). *Gesammelte Aufsätze zur Soziologie und Sozialpolitik*. Mohr.

Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, USA. <https://books.google.com?id=IFVJAbgySmEC>

