Incompatibility of emergence and flat ontology in assemblage theory

Valerii Shevchenko

2023

Table of Contents

Introduction

Scientific ontology describes objects of study for a given scientific discipline. Starting with "what do we study", we proceed to "how and with what methods can we gain reliable knowledge" about a given phenomenon. This algorithm is not different for social ontology. However, there are inherent difficulties with its application, for social ontology is not underpinned by any overarching principle like evolution in biology or lawful regularity expressible in mathematical form in economics.

...

Social ontology needs naturalistic constraints to more precisely define the objects of study, epistemology, and methodology of social science.

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 deriv-

atives. 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 mistakes.[1] 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 strategies.[2]

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 supposed to be a subject of a separate discipline, social science. Let

[1]There is an interesting similarity between a semantic regulatory mechanism like Harms' and regulatory networks in biology, that govern the dynamical repertoire of a given system like structural and regulatory genes [(Albert & Thakar, 2014)].

[2]An ESS is a strategy which, if adopted by a population, is resilient to invasion by any alternative strategy. Mathematically, an ESS can be defined as a strategy profile $s = (s_1, s_2, ..., s_n)$ such that $\forall s^{'} \neq s$, we have $\pi(s, s) > \pi(s, s^{'})$, where π is the average payoff of the population playing the strategies s and $s^{'}$ (Maynard Smith, 1982).

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 mechanism.[3]

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

[3]Replicator dynamics is a mathematical model used to describe the evolution of biological populations. It is based on the idea that individuals in a population can replicate themselves over multiple generations, and that their success or failure depends on their behavior relative to other members of the population. Mathematically, it is given by $\dot{x_i} = x_i(f_i(x) - \bar{f}(x))$, where x_i is the proportion of individuals in the population exhibiting a particular behavior, $f_i(x)$ is the fitness associated with that behavior, and $\bar{f}(x)$ is the average fitness in the population.

states, their physical realization and evolution.

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 com-

plementary 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.

	L	R
L	1,1	-1,-1
R	-1,-1	1,1

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.

	O	T
O	-1,-1	1,1
T	1,1	-1,-1

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 selfand 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 selfand 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 conventional.[4]

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 equilibria with normative force. With Hindriks, they propose a "uni-

[4]Harms (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).

fied 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". ^3ac825

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 (2015, 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 equlibrium,[5] 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

[5]Nash 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 equilibrium,[6] 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 (**smith1982?**). 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-

[6] Correlated 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 tha 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-equlibria? 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 constraints 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. ^fbc225

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 phenomena.[7]

[7]For 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 structures.[8] 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.

[8] In the preliminary understanding, a process is more ontologically primary that 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 component.[9]

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 mechanisms.[10] 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;

[9]Harms (2004) discusses what he calls "primitive content" in animal singaling, 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.

[10]See 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.

[11] Entrainment 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 withinand 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

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 (F. 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),[12] but

[12]There is an interesting similarity between a semantic regulatory mechanism like

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*[13]
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 for 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 eqiulibrium 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 equi-*

Harms' and regulatory networks in biology, that govern the dynamical repertoire of a given system like structural and regulatory genes [(Albert & Thakar, 2014)].

[13]An ESS is a strategy which, if adopted by a population, is resilient to invasion by any alternative strategy. Mathematically, an ESS can be defined as a strategy profile $s = (s_1, s_2, ..., s_n)$ such that $\forall s^{'} \neq s$, we have $\pi(s, s) > \pi(s, s^{'})$, where π is the average payoff of the population playing the strategies s and $s^{'}$ (Maynard Smith, 1982).

	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.

librium and *evolutionary stable strategy* (ESS) based on uncorrelated 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 conventions' 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:[14] 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.

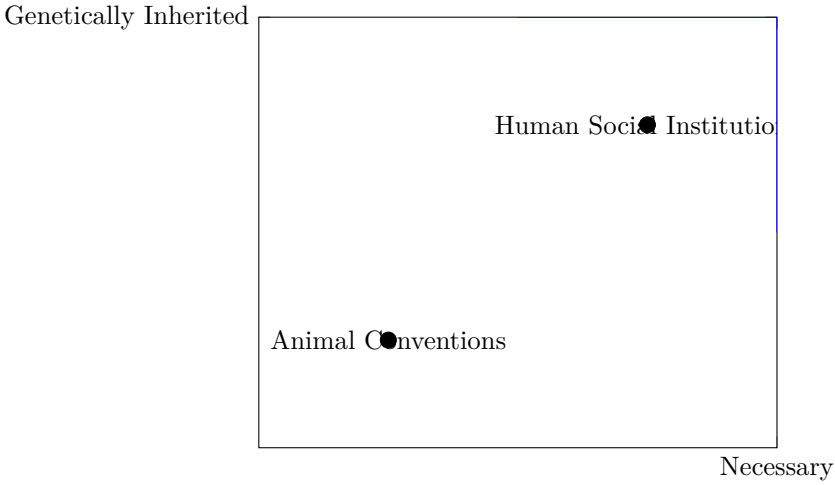
Furthermore, as Bicchieri (2005) asserts, conventions differ from social norms in their association with self-interest and common interest. While they converge with self-interest, they do not necessarily coincide with common interest. In the case of friends gathering at a bar, there

[14]Harms (2004) discusses what he calls "primitive content" in animal singaling, 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.

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



Essentially, social institutions are norm-driven conventions that require certain cognitive capacities which make recognition, complying to and changing of social norms possible.[15] Two 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?

Baraghith (2019) compares teleosemantic and game-theoretic views on emergence of conventions as public meaning. His main claim is that theories of Millikan (1987) and Skyrms (2010) share many aspects and can be synthesized to yield empirically testable and philosophically elaborated approach.

The author observes that signals, or public representations, become conventional by stabilization of a strategy profile in a Lewis signaling game, resulting in the emergence of behavioral regularities among involved agents. In other words, convention is generated by stabilization of a signaling system.

One of the similarities in teleosemantic and signaling approaches is that evolution drives the emergence of successful coordination between agents, be it parts of an organism or different organisms. However,

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

teleosemantic approach operates with the notion of a *function* (Millikan, 1987), whereas sender-receiver approach emphasizes *adaptive dynamics* by reinforcement learning (Skyrms, 2010).

In both approaches, conventions depend on their history and involve contingency. As Millikan (2005, p. 29) puts it:

> "A convention is merely a pattern of behavior that is (1) handed down from one person, pair, or group of persons to others – the pattern is reproduced – and (2) is such that, if *the pattern has a function*, then it is not the only pattern that might have served that function about as well. Thus, if a different precedent had been set instead, a different pattern of behavior would probably have been handed down instead."

As Baraghith notes, most criticism of teleosemantic view of the emergence of conventions has been that content—or representation of a world state by a sender—lacks explanation solely by its adaptive function or history. However, as Neander and Shea show, teleosemantics might solve the problem of mental content, intentionality, and thus, representation (Neander, 2008; Shea, 2018). In its turn, sender-receiver approach has received criticism for being atomistic and not able to accommodate "mental life"—cases with agents having advanced cognitive capacities like midredaing (Sterelny, 2017).

Baraghith stresses the crucial difference between speaker meaning and public meaning. In other words, a convention as a signaling system involves two kinds of information: a representation of an observed world state by a sender, and a signal sent from sender to receiver. The former is internal, and the latter is external. Representation and signal are mental and behavioral parts of a representational system, respectively.

A signaling game represents a coordination problem between world states, signals and acts, which are associated probabilistically. The most simple case has two states $W = \{\sigma_1, \sigma_2\}$, two messages $M = \{m_1, m_2\}$ that a sender S can transfer to a receiver R, and two acts $A = \{\alpha_1, \alpha_2\}$, by which R can respond to a received signal. There are pure sender and receiver strategies. The former is a function $s : W \mapsto M$ from world states to signals, and the latter is a function $r : M \mapsto A$ from signals to acts. With two signals and two acts, both sender and receiver have 4 strategies each. Assuming that all strategies

are equiprobable, 16 strategies are possible, from which only two are beneficial for both agents and constitute a *strict Nash equilibrium.*[16]

%%in latex, draw an extensive form signaling game with two world states, two signals and two acts. mark strict Nash equilibria with arrows%%

Lewis (2008) suggests that strict Nash equilibria can establish *signaling systems*, which become conventional in a population of senders and receivers.

In an evolutionary perspective of Skyrms (2010), signaling systems are not strict Nash equilibria, but *evolutionary stable strategies* (ESS). It is a strategy which, if adopted by a population, is resilient to invasion by any alternative strategy. It can be defined as a strtegy profile $s = (s_1, s_2, ..., s_n)$ such that $\forall s^{'} \neq s$, we have $U(s, s) > U(s, s^{'})$, where U is the average payoff of the population playing the strategies s and $s^{'}$ (Maynard Smith, 1982). On this account, given an adaptive process that guides the behavior of agents, any signaling game iterated over time results in an ESS. Depending on initial conditions, population converges on one of the two signaling systems, what is often modeled with replicator dynamics.[17]

Another detail of this approach is its connection to information theory. A signal m_1 carries information if it changes the probabilities of any world state. The information quantity is measured by how far the probability is moved, and information content—by direction of probability: increasing or decreasing. Franke and Wagner (2014) show the Bayesian likelihood of a world state σ_i given a signal m_j:

$$P(\sigma_i \mid m_j) = \frac{P(m_j \mid \sigma_i) \times P(\sigma_i)}{\sum_t P(m_j \mid \sigma_i) \times P(\sigma_i)}$$

It means that if state σ_i occurs with prior probabilities $P(\sigma_i)$ and

[16]There is an interesting similarity between a semantic regulatory mechanism like Harms' and regulatory networks in biology, that govern the dynamical repertoire of a given system like structural and regulatory genes [(Albert & Thakar, 2014)].

[17]Replicator dynamics is a mathematical model used to describe the evolution of biological populations. It is based on the idea that individuals in a population can replicate themselves over multiple generations, and that their success or failure depends on their behavior relative to other members of the population. Mathematically, it is given by $\dot{x_i} = x_i(f_i(x) - \bar{f}(x))$, where x_i is the proportion of individuals in the population exhibiting a particular behavior, $f_i(x)$ is the fitness associated with that behavior, and $\bar{f}(x)$ is the average fitness in the population.

$P(m_j \mid \sigma_i) > 0$, a signal m_j is sent. Signals may initially contain no intrinsic meaning, and the dynamics does not require any sophisticated cognitive capacities of the agents. They do not need to have pre-existing mental language for a signaling system to be established (Skyrms, 2010, p. 7). This makes sense in "animal conventions", but not easily so in human ones. Thus, according to Skyrms' sender-receiver approach, convention is an ESS. It makes a lot of sense in "animal conventions", but its application to human social institutions is not straightforward. As Huttegger puts it (2007, p. 413):

> "There is at least one functional aspect of human languages that can fundamentally be expressed in terms of signaling systems: communication facilitates social coordination".

However, it is not sufficient for evolutionary account of human social coordination resulting in social institutions.

Another important formal approach to the emergence of conventions is due to Harms (2004). He synthesizes sender-receiver framework and Millikan's teleosemantics. According to this approach, any semantic convention, or "rule", might be considered as a "function-stabilizing mechanism". It helps to coordinate the behavior of different organisms or different parts of an organism to perform an evolutionary adapted biological function. Rules are sets of maps from conditions to processes one by one. They say what to happen next given a state of the world. Rules for evolutionary adapted traits (AT) might be expressed as

$$R_{AT} = \{\langle c_i, p_i \rangle \mid ATselp_i inc_i\}$$

A rule for an adaptive trait is a set of all ordered pairs of a condition and a process such that the trait was selected for performing the process p_i in the conditions c_i. (Harms, 2004, p. 203).

It has been observed that animal signals not only inform about the world states, but also direct the behavior of others. For example, alarm calls of vervet monkeys both convey "Look, there is a leopard!" and "Run up the nearest tree" (Baraghith, 2019; Seyfarth & Cheney, 1990). Harms calls this "primitive content" that has both indicative and imperative functions (Harms, 2004, p. 189). Millikan calls it "pushmi-pullyu" representation and notes that purely descriptive and directive representations require a more advanced cognitive process than primitives (Millikan, 2005, p. 166).

Evolutionary development of primitive content leads to the divergence of its descriptive and directive functions due to advanced cognitive capacities. As Harms suggest, it introduces a stabilizing, or regulatory mechanism SM that works "atop" of conventions as rules for adaptive traits and guides behavior in case of failure of R_{AT}. It employs a corrective signal $CS = \{cs_1, ..., cs_n\}$ to "enforce" the initial convention when a signal is not sent in the presence of a world state it was selected for:

$$R_{SM} = \{\langle \sigma_i \wedge \neg m_j where \langle \sigma_i, m_j \rangle \in R_{AT} \rangle \mid SM selcs when(\sigma_i \wedge \neg m_j)\}$$

The rule for a stabilizing mechanism is a set of ordered pairs consisting of the failure of an adaptive trait in the first place and a corresponding corrective signal in the second place. If the adaptive trait fails, the stabilizing mechanism will detect this failure and send a corrective signal/action to restore it.[18] This division is echoed in Millikan's work as firstand higher-order reproductive families (Millikan, 1987, p. 23). According to it, conventions R_{AT} are firstand stabilizing mechanisms R_{SM} are second-order reproductive families that serve the same goal of restoring a first-order proper function.

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?

F. 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

[18]There is an interesting similarity between a semantic regulatory mechanism like Harms' and regulatory networks in biology, that govern the dynamical repertoire of a given system like structural and regulatory genes [(Albert & Thakar, 2014)].

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

Incompatibility of emergence and flat ontology in assemblage theory

Introduction

References

Albert, R., & Thakar, J. (2014). Boolean modeling: A logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *WIREs Systems Biology and Medicine*, *6*(5), 353–369. https://doi.org/10.1002/wsbm.1273

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

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 Teleosemantic and Game Theoretical Perspective. *Journal of Philosophy*, 30.

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/sum2018/entries/social-ontology/

Ferguson, A. (1980). *An Essay on the History of Civil Society, 1767*. Transaction Publishers.

Franke, M., & Wagner, E. O. (2014). Game Theory and the Evolution of Meaning: Game Theory and the Evolution of Meaning. *Language and Linguistics Compass*, *8*(9), 359–372. https://doi.org/10.1111/lnc3.12086

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

Glennan, S. (2017). *The New Mechanical Philosophy* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198779711.001.0001

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=ztl99e9ugtAC

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. 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

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

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., & Hyyryläinen, 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

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

Millikan, R. G. (1987). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press. https://books.google.com?id=jncHBAYe8TkC

Millikan, R. G. (2005). *Language: A biological model*. Clarendon Press ; Oxford University Press.

Neander, K. (2008). Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality? In M. Ruse (Ed.), *The Oxford Handbook of Philosophy of Biology* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195182057.003.0017

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

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

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?id=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

Seyfarth, R., & Cheney, D. (1990). The assessment by vervet monkeys of their own and another species' alarm calls. *Animal Behaviour*, *40*(4), 754–764. https://doi.org/10.1016/S0003-3472(05)80704-3

Shea, N. (2018). *Representation in cognitive science* (First edition). Oxford University Press.

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.

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.

Sterelny, K. (2017). From code to speaker meaning. *Biology & Philosophy*, *32*(6), 819–838. https://doi.org/10.1007/s10539-017-9597-8

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). Ra-

tionality 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

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