Coordination as naturalistic social ontology: constraints and explanation

Valerii Shevchenko

Abstract

In the paper, I propose a project of naturalistic social ontology based on rules-in-equilibria theory of social institutions (Guala & Hindriks, 2015; F. Hindriks & Guala, 2015). On this account, institutions are correlated equilibria with normative force. However, if both humans and animals solve coordination problems in a structurally similar way, and only humans have social institutions, how the latter arise evolutionary? I suggest identifying possible causes among cognitive capacities like mind-reading and imitation by building dynamic models. In conclusion, the relationship between involved types of explanation, mechanistic and equilibrium ones, is discussed, as well as implications for social ontology.

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

Introduction

Social ontology studies "what there is" in the social world. According to Epstein (2018), this study divides into two distinct inquiries. The first one deals with the constituents of social entities and addresses the question "what the social world is 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—the latter (Epstein, 2015, p. 57).

The difference between the strands is in metaphysical relation between individuals and social facts: in the first case, social facts supervene on the facts about individuals, meaning that social properties cannot change without changing of individual ones. And in the second case, social facts supervene on the facts about material objects like dollar bills for money and a line of rocks for borders. 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[1].

What does it mean for social ontology to be "naturalistic"? Broadly conceived, it means striving to "cut the nature of the social world at its joints". But where are these in terms of grounds and anchors? I take it to be naturalism about anchors, or mechanism of bringing

[1] The Searlian 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.

instantiation conditions about. On this account, naturalistic social ontology addresses natural constraints of and their influence on social phenomena, namely, the role of representations, mental states, their physical realization and evolution.

In the current paper, I describe the constraints for a naturalistic extension of the rules-in-equilibria theory of social institutions put forward by Guala and Hindriks as a unified social ontology (Guala & Hindriks, 2015; F. Hindriks & Guala, 2015). It criticizes and builds upon Searle's theory by arguing that instantiation conditions for social kinds are brought about by social coordination. 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. My extension consists of addressing the evolutionary constraints on equilibrium emergence and necessary and sufficient cognitive capacities for it. In other words, I address the question "if both humans and animals solve coordination problems, and only humans have social institutions as correlated strategies with inherent deontic, or normative, powers, how the latter arise evolutionary and what cognitive capacities are needed for it?". Answering this question would provide a naturalistic basis for scientifically informed social ontology, and I sketch a possible way to do it.

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 and called rules-in-equilibrium theory of social institutions (Guala, 2016; 2015; 2015). In the second section, I describe naturalistic constraints for the theory of coordination that include cognitive and evolutionary requirements for emergence and persistence of institutions. The idea here is that if cognitive capacities have been involved in emergence of social institutions, explicating them would bring social ontology closer to non-reductive physicalism about social entities. This position suggests ontological, but not theoretical reduction of social entities to cognitive ones. I conclude the paper discussing the form of explanation involved in the proposed extension. It is relevant, for the formulation of the issue implies a deeper concern about integrating causal-mechanistic explanation widespread in cognitive science and evolutionary theory with an equilibrium, or structural explanation, found in game theory.

Game theory as a model for social ontology

The tradition of treating coordination as basic for the social world is both historically rich and thriving. It dates back to Hobbes (2016, p. 1651) with his proto-game-theoretic treatment of rational deliberation of individual agents as a source of jointly optimal social decision, an agreement to form a state. The so-called "Hobbesian problem" of social order emergence has been popularized and given the status of the main theoretical problem of sociology by Parsons (1937). 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 convened on 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 fundamental part of human social order and ontology have been discussed long before the rise of game theory, and it became especially convenient to express this issue with it. Famously, Skyrms (2014), treated the "Hobbesian problem" of social contract emergence in explicitly game-theoretic terms of evolutionary dynamics of deliberation.

However, 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 emergent 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 unless everybody does the same. As conventions are defined in terms of coordination problems, it is useful to say about them as well.

Coordination problems are situations where agents have common interests and it is not evident how exactly they can be met. O'Connor (2019) distinguishes two classes of coordination problems, correlative and complementary ones. Correlative coordination problems are problems where 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 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 different sides, they will stay safe. One important feature of this and other coordination problems is arbitrariness, meaning that it does not really 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	$1,1$	$0,0$
R	$0,0$	$1,1$

Figure 1: Driving game (correlative coordination problem)

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 quests come, and if they both order pizza delivery, they will have plenty of food but be embarrassed by the mess at the house.

	T	O
T	$0,0$	$1,1$
O	$1,1$	$0,0$

Figure 2: Party game (complementary coordination problem)

Both correlative and complementary coordination problems can lead to emergence of social institutions and, hence, are important to social ontology.

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, or deontic powers. 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 friends. On the contrary, if two cars are driving on the same side of the road towards each other, the drivers are enforced to swerve, for otherwise they might crash. They ought to swerve, for not only they might be fined but they might cause an accident. However, the relationship between conventions and social norms is not clear cut. Bicchieri et al. (2018) distinguish between them on the basis of the relationship between self- and common interest. In norms, self-interest might contradict common interest, meaning that one might overcome herself when complying to a social norm, whereas there is no such tension in conventions. However, Guala (2013) shows that norms and conventions are continuous, for normativity arises as an intrinsic feature of conventions. O'Connor underlines the idea that conventions and norms are poles of a continuum along which the former acquire normative force.

The second distinction concerns 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 in bacteria or vervet monkeys might well be modelled as conventions in Lewisian sense of repeated behavioral patterns of solving coordination problems (cf. Harms, 2004; Skyrms, 2010), but it is not historically contingent in the sense of several possible solutions being equally profitable, for there is evolutionary constraints responsible for this. This distinction, as O'Connor underlines, illuminates some conventions as more functional and other as more conventional.

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, they 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 in what relation do institutions stay to conventions and norms.

Guala (2016) offers a definition of institutions as sets of correlated strategies with deontic powers, or normative forces. 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). It bridges accounts of regulative rules of the form "if X, do Y" that agents employ in complying to norms, equlibria of strategic games and constitutive rules of the form "X counts as Y in C", where the former two are complementary and comprise a rules-in-equilibria account, and the latter supplements it by providing a symbolic representation.

The rule-based account conceives of social institutions as rules guiding and constraining behavior in social interaction or "humanly devised constrains" of social interactions. In sociology, the tradition of treating institutions as rules dates back to classical figures like Weber (1924) and Parsons (2015, p. 1935). The equlibrium-based account sees institutions as behavioral regularities and, most importantly, solutions to coordination problems. The constitutive rules account, introduced by Searle (2010, 1995), sees institutions as systems assigning statuses and functions to physical entities.

According to the authors, 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. Hindriks and Guala illustrate this point by comparing the two paradigmatic games from game theory, which are prisoner's dilemma and stag hunt.

Although mutual defection in the prisoner's dilemma is a Nash equlibrium[2], it is not a social

[2]Nash equilibrium is a solution concept describing a strategy profile consisting of each player's best

	C	D
C	$-1, -1$	$-3, 0$
D	$0, -3$	$-2, -2$

	S	H
S	$4, 4$	$1, 3$
H	$3, 1$	$2, 2$

Figure 3: Prisoner's dilemma (left) and Stag hunt (right)

institution, for it is not self-sustaining due to 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 correlation of players' strategies to achieve a bigger joint payoff. The latter means that the strategy is salient and beneficial for players, what explains why some rules are followed and others not.

However, the notion of players' correlated strategies, or correlated equilibrium[3], as an explanans of the stability of institutions is insufficient too, as the authors point out, for it is too loose. The authors provide an example of non-human animals solving coordination problems with correlation devices, but still not having institutions. For example, male baboons, lions, swallowtails and some 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 minimizes possible damage to both parties and lets the incumbent 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 here are only one number in each cell.

	H	D	B
H	-1	2	0.5
D	0	1	0.5
B	-0.5	1.5	1.0

Figure 4: Hawk-dove-bourgeous game

Maynard Smith calls this solution "bourgeous equilibrium", for it is an implicative strategy of the form "Hawk if owner" and "Dove if intruder". O'Connor (2019) points out that this strategy profile is based not on correlation, but on uncorrelated asymmetry. It is a feature of games where players extract additional bits of information from environment and which are not included in the structure of a game. For example, they know that they are "Player 1" and "Player 2" rather than their roles are randomized. This underlies an important methodological distinction between different solution concepts: correlated equilibrium and evolutionary stability[4]. However, as Skyrms points out, "bourgeous" is correlated equilibrium spontaneously arising from symmetry-breaking that happens when individuals randomize the choice of their strategies and do not know whether they are "hawkes" or "doves" (2014, p. 78).

Guala and Hindriks argue that human and animal conventions differ only in the scope of actionable signals. As they formulate it, the set of signals to which animals might respond is rather narrower than that of humans, due to tight coupling of stimulus and response to achieve coordination. 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

response to the other player's strategies where no one gains bigger payoff by deviating unilaterally.

[3]Correlated equilibrium is a more general solution concept than Nash equilibrium introduced by Aumann (1987, 1974). 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.

[4]Evolutionary stable is a solution concept viewing relation between population of the same species. It is a strategy profile resistant to strategies of "mutants", which is a best response to itself.

representation of the environment conditioning behavior. In other words, humans are able to invent and follow different rules given the same correlation device. 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.

However, it is not evident how exactly rules represent strategies. To clarify this issue, the authors, drawing on F. A. Hindriks (2005), propose to bridge their rules-in-equilbria 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.

However, the are several problems with this approach. First, regarding the project of naturalistic social ontology, game theory overall is not intrinsically naturalistic. It is a tool that models situations of interdependent decision-making, the process of which is not by itself inherently connected to naturalism. The only connection is the use of game theory to model evolutionary processes (cf. Maynard Smith, 1982). Next, there have been doubts about empirical adequacy of game theory (see Guala, 2006).

The unified rules-in-equlibria account has several shortcomings, as well. Although it mentions that social coordination as correlation of strategies with deontic powers seems to have evolutionary roots, this approach does not address cognitive and evolutionary constraints that shape and make coordination in the form of social institutions possible. However, Guala (2020), one of the authors of the unified social ontology, is concerned with a similar issue and asks, "what cognitive mechanisms establish coordination?". Moreover, some other accounts of coordination mentioned above make some reference to cognitive mechanisms helping to make a decision in converging on a certain equilibrium. For example, Bicchieri (2005; 2018) refers to social norms as being "activated" by scripts and grounded in cognitive schemata. However, she does not explain mechanisms of this activation.

The main issue with the game-theoretic approach to social coordination, as Kaidesoja et al. (2019) argue, is that explanations in terms of equilibria do not explicate any causal processes or mechanisms. Guala himself recognizes this feat, for it is applicable to the rules-in-equlibria account as well. Hence, a naturalistic explanation of emergence and persistence of social institutions as rooted in evolution needs to be supplemented with cognitive mechanisms that enable individuals to coordinate. But what these mechanisms might be?

Naturalistic constraints on coordination: cognition, evolution and physical realization

To discuss cognitive mechanisms in relation to social coordination, the issue of the connection between social and cognitive phenomena needs to be addressed first. S. Turner (2007) argues that social theory relies heavily on and makes generous use of mentalistic and cognitive concepts like "meaning", "belief", "collective representations" and others. However, at the same time it employs the strategy of minimizing the cognitive by using these concepts only metaphorically and not engaging with the actual cognitive science. This applies to social ontology, as well, at least to approaches within the "Standard model" that employ the notions of intentionality, belief, acceptance and others.

Turner proposes what he calls a "sane" constraint on social theory which might be extended to social ontology as it is the most abstract edge of theory as discussed earlier. According to this constraint, both theory and ontology ought to be physically, computationally and cognitively realistic. It means that its concepts and explanations must be grounded in physically realizable cognitive entities and be computationally tractable. And if game-theoretic explanations of coordination are computationally tractable, the other two requirements remain unsatisfied.

S. P. Turner (2018), drawing on Guala (2016), who discusses the possibility of attaining an equilibrium without explicit prior agreement with the help of a physical marker that he calls a "focal point", 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 certain direction. In other words, coordination is the main mechanism of scaling up from autonomous agents to larger social groups and societies, which is itself grounded in cognitive and neural phenomena like joint attention and mirroring.

However, how to untangle the relationship between these domains regarding coordination as the basic mechanism of social ontology and ask meaningful questions about it? Two types of questions might be asked.

First, if normative force is what distinguishes coordination problems in animals and humans, how does it emerge? And what are minimal cognitive requirements that lead to it? Methodologically, it is a game-theoretic problem of evolutionary transformation of a coordination game with additional cognitive capacity like mindreading, imitation, teleological thinking or others. 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 contribute to the emergence and persistence of social institutions, and this raises another set of questions.

The second set of questions is more general and concerns the relationship between social and cognitive phenomena. If the latter ones are admitted to have explanatory and ontological significance in social phenomena, as Turner and Guala independently propose, two positions are possible. Strydom (2007) formulates them as strong and weak cognitivism.

Strong cognitivism presupposes the ontological priority of cognitive and neurophysiological entities and processes over social ones like beliefs and intentions and describes them in cognitive, informational, cybernetic and biological terms. Methodologically, strong cognitivism subscribes to Sperber's infra-individualism which refers to an organism's productive processes lying below the level of individual's intentions stretching from brain modules to genes. The two major versions of strong cognitivism regarding social science is Sperber's (1997) and Turner's (2002) ones. Turner comes close to an eliminativist position disregarding all the versions of social theory built on traditional social scientific concepts. Sperber's epidemiology of representations approach (2011), however, argues for ontological reduction without theoretical one. Hence, in strong cognitivism, there are at least two possible strategies: eliminativism and reductionism about social entities.

Weak cognitivism, in its turn, presupposes ontological autonomy of the social and describes it in terms of beliefs, intentions, rational choice and other "mentalistic" concepts, which Turner criticizes. Methodologically, weak cognitivism subscribes to individualism which comes in two flavors, weak and strong ones. The former treats an individual as a basic unit of social structure, whereas the latter treats and individual as a coherent conscious subject having "full-fledged" reasons and performing actions. This type of individualism is opposed to Sperber's infra-individualism. Rules-in-equilibria does not subscribe to either form of individualism.

Thus, if one addresses the problem that game-theoretic explanations of coordination do not explicate cognitive mechanisms responsible for it, she stumbles upon the issue of the relation between social and cognitive entities. There are at least three options. First, it is non-reductive physicalism like Sperber's that allows ontological reduction without theoretical one. On this account, social institutions as rules-in-equilibria might be ontologically reduced to cognitive mechanisms responsible for representations of these institutions and for reflexive loops. However, it also retains explanatory autonomy of the game-theoretic notions of coordination. Second, it is reductive physicalism like Turner's that presupposes eliminativism about social scientific concepts. On this account, naturalistic social ontology is not "social", but consists only of neurophysiological states of agents and retains instrumental use of game-theoretic notions of coordination, as Turner does in his (2018). Third, it is a more classic non-physicalist option retaining ontological autonomy of social entities employing the notions of beliefs, representations and mindreading but not addressing their natural constraints. It is a kind of realism regarding social kinds that Guala advocates while sidestepping individualism.

If one inquires into the role of evolved cognitive capacities in emergence of equilibria with normative force, involvement of these capacities, however, does not entail ontological reducibility of social institutions to cognitive capacities like Sperber and Turner propose, but it assumes their strong influence. If social institutions are correlated strategies with deontic powers, understood as behavioral regularities in populations, showing involvement of cognitive capacities in emergence of these behavioral regularities would be a proof-of-concept of their close relation to social ontology.

There are two questions to be addressed afterwards. First, an empirical one about concrete cognitive mechanisms involved into coordination problems in experimental settings. Testing intuitions like most social coordination in everyday life is belief-less, meaning that it does not involve explicit propositional attitudes that cause behavior (Guala, 2020) is an example of such question. To address this problem, Sarkia et al. (2020) propose a framework illuminating the division of labor between cognitive and social scientists studying social phenomena. It is based on mechanistic philosophy of science (Craver & Darden, 2013; 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 it consists 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 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. And this, supposedly, is where game-theoretic explanations of coordination come in. The second question that might be asked afterwards is a theoretical one. It is about bridging extended rules-in-equilibria with these accounts of coordination on a micro-level. It means inquiring into emergence and persistence of social order in small social groups as opposed to populations. There are models connecting these levels in a hierarchical way, but that are focused on linguistic conventions instead of social (Hawkins et al., 2021).

Answering these questions would bring social ontology closer to non-reductive physicalism in Sperber's style, that allows ontological reduction without theoretical one. It would mean that sociological concepts like roles, from which the notion of social structure emerges are ontologically dependent on social institutions which presumably emerge with the help of evolved cognitive capacities. However, it involves a harder question of relating two different types of explanation: equilbrium, or structural one, and mechanistic, or causal one, with which I conclude the paper.

Conclusion: equilibrium and mechanistic explanations

Asking the questions about cognitive mechanisms responsible for social coordination as presented by equilibria is inherently problematic in conceptual and semantic senses. Conceptually, because equilibrium, or structural explanations are in tension with causal, or mechanistic ones, for, as Sober (1983) points out, the former show how an event might have occurred regardless the actual causes involved, whereas the latter shows how an event was actually produced. They differ in the source of their explanatory force: mechanistic explanations show entities and their relations as responsible for a phenomenon, whereas equilibrium explanations 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, as well, for it is not evident what precisely does the notion of a mechanism mean regarding naturalistic explanation of coordination as social ontology. Is coordination itself a mechanism? Let us start from the latter issue.

The notion of coordination has different meanings in two relevant vocabularies, 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, meaning that all agents get highest possible payoff in a coordination game. In the latter case, coordination (C2) is a also a successful result of social interaction, either spatiotemporally synchronous or not, that is possible due to 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). These capacities are themselves 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) [5].

Despite the similarity of both meanings of coordination being successful results of some processes, their underlying presuppositions are drastically different. 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, 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 mathematical rather than empirical structure of coordination. The problem, then, is relating mathematical and empirical structures of the presumably same phenomenon. For the current research, it is not relevant yet, for intermediate steps must be undertaken. But a similar explanatory problem, that of causality of equlibrium explanation itself, arises.

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 system’s initial conditions and dynamic processes as

[5] 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).

possible causes understood as interventions, meaning variables subject to manipulation and control on which an outcome depends (Woodward, 2005). Discussing 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 on.

It follows, then, that equilibrium explanations that use manipulation of parameters, are causal explanations. Despite being causal, however, they are not mechanistic, for they still do not provide an "entities and relations" picture. Inquiring into cognitive capacities that might have affected emergence of equilibria with deontic powers 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.

In sum, naturalistic social ontology can be arrived at methodologically with inference from game theory, for it is the only language naturalistic extensions of which do not change its substantial meaning and explanatory power. It is realist and fallibilist, for it strives to describe the relevant aspects of the world's structure and provides models that can be tested against empirical data. Guala's and Hindrik's unified social ontology known as rules-in-equlibria, treats social institutions as objective results of correlation between agents' strategies and their use of social norms as explicit or implicit rules of conduct. Although the authors note that this correlation is similar to that of bourgeous equilibrium in animals, the transition from the latter to the former is obscure, for it is not evident how do deontic powers of certain strategies arise, leading to social institutions. My proposition is to search for possible sources of this normativity in human evolutionary history by modeling major cognitive capacities that are usually said to be uniquely human, for example, mindreading, imitation and others. This would bring social ontology closer to non-reductive physicalism that allows ontological reduction without theoretical one.

References

Aoki, M. (2007). Endogenizing institutions and institutional changes*. *Journal of Institutional Economics*, *3*(1), 1–31. https://doi.org/10.1017/S1744137406000531

Aumann, R. J. (1987). Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, *55*(1), 1. https://doi.org/10.2307/1911154

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

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.

Gilbert, M. (1992). *On Social Facts*. Princeton University Press. https://books.google.com?id=yYvcDwAAQBAJ

Glennan, S. (2017). *The New Mechanical Philosophy*. Oxford University Press.

Guala, F. (2013). The normativity of Lewis Conventions. *Synthese*, *190*(15), 3107–3122. https://doi.org/10.1007/s11229-012-0131-x

Guala, F. (2006). Has Game Theory Been Refuted? *The Journal of Philosophy*, *103*(5), 239–263. https://www.jstor.org/stable/20619937

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

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). *From partners to populations: A hierarchical Bayesian account of coordination and convention*. http://arxiv.org/abs/2104.05857

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

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

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

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.

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. (2015). The Place of Ultimate Values in Sociological Theory. *The International Journal of Ethics*. https://doi.org/10.1086/intejethi.45.3.2378271

Parsons, T. (1937). *The Structure of Social Action*. The Free Press.

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. (2010). *Making the Social World: The Structure of Human Civilization*. Oxford University Press. https://books.google.com?id=kz6R0eDZ5OEC

Searle, J. (1995). *The Construction of Social Reality*. Simon and Schuster. https://books.google.com?id=zrLQwJCcoOsC

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. (1997). Individualisme méthodologique et cognitivisme. In R. Boudon, A. Bouvier, & F. Chazel, *Cognition et sciences sociales* (p. 121). Presses Universitaires de France. https://doi.org/10.3917/puf.boudo.1997.01.0123

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.

Strydom, P. (2007). Introduction: A Cartography of Contemporary Cognitive Social Theory. *European Journal of Social Theory*, *10*(3), 339–356. https://doi.org/10.1177/1368431007080699

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

Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford University Press. https://books.google.com?id=6ltpAgAAQBAJ

Turner, S. (2002). *Brains/Practices/Relativism: Social Theory After Cognitive Science*. University of Chicago Press. https://books.google.com?id=4sYG0Fj9Y1MC

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. P. (2018). *Cognitive Science and the Social: A Primer* (1st ed.). Routledge. https://doi.org/10.4324/9781351180528

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