Bourgeois equilibrium in rules-in-equilibria theory

Valerii Shevchenko

Abstract

In this paper, I argue that the emergence of evolutionary stable correlation is the core issue of naturalistic social ontology. According to the rules-in-equilibria theory, social institutions are the central unit of social ontology [@guala2016b], and coordination is its main mechanism rooted in evolution [@shevchenko2023]. As institutions are normatively-driven self-sustaining behavioral regularities designed to solve coordination problems [@lewis2008; @aoki2007], they share many features with 'animal conventions' that help animals solve coordination problems and maintain stable relationships [@hindriks2015]. 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.

Contents

Introduction

In this paper, I argue that the emergence of evolutionary stable correlation is the core issue of naturalistic social ontology. According to the rules-in-equilibria theory, social institutions are the central unit of social ontology [@guala2016b], and coordination is its main mechanism rooted in evolution [@shevchenko2023]. As institutions are normatively-driven self-sustaining behavioral regularities designed to solve coordination problems [@lewis2008; @aoki2007], they share many features with 'animal conventions' that help animals solve coordination problems and maintain stable relationships [@hindriks2015]. 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 [-@guala2016b] 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 [-@searle1995],[1] but by regulative rules of the form "do X if Y"
2. from a game-theoretic point of view, regulative rules can be seen as agents' strategies that comprise a *correlated equilibrium*[2]
3. constitutive rules are linguistically transformed regulative rules with added theoretical term that represents a certain equilibrium

[1] Coordination equilibrium is a concept defined by philosopher David Lewis which states that when two or more individuals are engaged in a coordination game, they will naturally gravitate towards the same outcome, as this is the most rational choice. The idea is that each individual will tend to choose the same outcome because they can both benefit from it. This is in contrast with Nash equilibrium, where each individual must make a choice that maximizes their own payoff without considering the other's payoff.

[2] Correlated equilibrium is a general solution concept introduced by Aumann [-@aumann1974; -@aumann1987]. As opposed to the classic Nash equilibrium, where players choose their strategies independently, here players choose strategies based on a public signal the value of which they assess privately, thus coordinating their actions according to a given correlation device.

4. at the same time, many animal species including baboons, lions, swallowtails, and others exhibit behavioral patterns describable in the form similar to correlated equilibrium [@maynardsmith1982]
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 [-@sterelny2003], 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. For this theory to be naturalistic, there should be computationally tractable and plausible mechanisms of transition from 'animal conventions' to social institutions, which are now lacking. And the relation of biological and cultural evolution involved in this transition should be clarified, as well.

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

	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.

Guala and Hindriks interpret bourgeois equilibrium as a correlated one. However, there are at least two interpretations of it: *correlated equilibrium* (CE) and *evolutionary stable strategy* (ESS)[3] 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 CE, and there comes the need for showing how the former becomes the latter in the course of evolution;
- or that 'animal conventions' are themselves correlated, and there comes the need for 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 correlated equilibrium in game theory [@skyrms1994; @lee-penagos2016; @kim2017; @metzger2018; @herrmann2021], 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 [@zawidzki2013], 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 @shevchenko2023.

This paper is structured as follows. The first section is devoted to description of the rules-in-equilibria (RiE) theory of social institutions. In the second section, the relationship between social institutions, conventions,

[3]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 $s^{'}$ [@maynardsmith1982].

and norms is discussed. The third section examines the notion of representation as used in RiE. In the fourth section, the role of Hawk-Dove-Bourgeois game as correlation and as asymmetry of strategies is studied. The final section explores the source of randomization in correlation as the problem in social institutions as evolved correlated equilibria. Let us start with the notion of social institutions as rules-in-equilibria.

1. Social institutions as rules-in-equilibria

@hindriks2015 present a unified theory of social institutions as rules in equilibria represented symbolically by theoretical terms like "money" or "marriage". It bridges accounts of regulative rules, equlibria of strategic games and constitutive rules, where the former two are complementary and comprise a rules-in-equilibria account, and the latter supplements it by providing a symbolic representation.

According to @guala2016b, institutions as rules-in-equilibria are normatively-driven behavioral regularities comprising 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" [@north1990], "norm-governed social practices" [@tuomela2013] and "self-sustaining salient behavioral patterns" [@aoki2007]. 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.

The rule-based account conceives of social institutions as rules guiding and constraining behavior in social interaction or "humanly devised constraints" of social interactions [@north1990]. In sociology, the tradition of treating institutions as rules dates back to such classical figures as @weber1924 and @parsons2015, and it continues to thrive today. The equlibrium-based account sees institutions as behavioral regularities and, most importantly, solutions to coordination problems. The constitutive rules account sees institutions as systems assigning statuses and functions to physical entities [@searle1995].

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 (NE),[4] it is not a social institution, however, 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 NE, is 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 other not.

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

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

Table 2: Prisoner's dilemma (left) and Stag hunt (right)

However, the notion of players' correlated strategies as an *explanans* of the stability of institutions is insufficient, as the authors point out, for it is too permissive. The authors provide an example of non-human animals solving coordination problems 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. 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 [@maynardsmith1982]. The authors use Maynard Smith's exposition of animal territorial behavior represented as a "Hawk-Dove-Bourgeois" game to provide an example of a prototypical social institution:

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

Table 3: small "Hawk-Dove-Bourgeois" game

[4] 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.

Presented with a terrestrial resource, a "Hawk" player fights over it, a "Dove" retreats and "Bourgeois" uses a strategy "fight if own and retreat if do not own". In this game, Guala and Hindriks see the "bourgeois" strategy "fight if own" as a correlated one, meaning that players coordinate their actions by conditioning them on an external signal. As the author point out, it is a "simple pre-emption device: whoever occupied the land first has the right to use it" [@hindriks2015, 465]. The temporal order of occupation is used as a correlationon device. Overall, this correlation fulfills the necessary condition of being an institution.

@guala2015 illustrate the applicability of HDB to humans with a game where two tribes, spatially separated by a dry river, graze their cattle. The dry river serves as a "focal point"—a salient feature of the environment that the members of both tribes have been aware of [@schelling1980]. It also serves as a correlation device, for it is a source of a public signal that coordinated actions of different tribes without their explicit agreement. Thus, the shepherds of both tribes have three possible strategies: "Graze", "Not graze" and "Graze if North / South of the river" according to the history of their territorial occupation. The members of one tribe might be killed by the members of another if grazing their cattle on another side of the dry river which the other tribe possesses. The most stable set of strategies is grazing if on the own side of the dry river. However, this is insufficient, for the payoff structure of the game is uniform for animal and human cases. Hence, we cannot discriminate between them solely on this basis.

	G	NG	GIS
G	-1	2	0.5
NG	0	1	0.5
GIN	-0.5	1.5	1.0

Table 4: Grazing game: the player strategies are Graze, Not Graze and Graze if North / Graze if South

As both cases are identical, what, then, distinguishes 'animal conventions' from human social institutions? Guala and Hidriks argue that it is the scope of actionable signals. Building on the work of @sterelny2003, they say that animals may only respond to a limited set of stimuli. However, humans are able to use representations and symbols to condition behavior, decouple stimulus and response and invent new rules. For example, butterflies cannot coordinate on anything but who occupied the sunspot first and unable to create new equilibria. Humans, however, can. The question is by what means this difference is reflected in RiE? Let us turn to the relationship between the concepts of rules, norms, conventions, and institutions used in RiE.

2. Rules, norms and conventions

Guala and Hindriks argue that an adequate theory of institutions must have three explanatory components [@guala2015, 469]:

- coordination
- correlation
- representation.

Following the logic of the authors, institutions coordinate behavior by correlation of strategies, and humans are able to devise many strategies and equilibria given the same correlation device, or signal, due to an advanced cognitive capacity for conditioning behavior on representation of the environment. At the same time, as RiE has "rules" and "equilibria" parts, they must be connected. For this reason, rules are symbolic representations of strategies in a game that comprise equilibria. They not only serve as symbolic markers of the properties of equilibria, but considerably save cognitive effort. As Aoki notes [-@aoki2007, 6]:

> "An institution is a self-sustaining, salient pattern of social interaction, as represented by meaningful rules that every agent knows, and incorporated as agents' shared beliefs about the ways the game is to be played".

```
graph LR;
subgraph Objective
3(Observer rules)
1(Correlated equilibrium)
1<-.->3
end
subgraph Subjective
3<-.Representation.->4
end
4(Agent rules)
```

Thus, rules are essentially social norms which agents formulate for themselves and adhere to. And strategies are an "objective" counterpart represented by these rules. However, it is not evident how rules represent strategies. To clarify, the authors, drawing on @hindriks2005, bridge RiE account of institutions with the constitutive rules account. The latter sees institutions as systems of statuses and functions, paradigmatically proposed by Searle [-@searle1995] 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 "do X if Y".

The constitutive rules approach argues that our beliefs are essential for the existence of institutions, which involve more than just actions. This applies to objects, persons, and events too—for example, "Bills issued by the Bureau of Engraving and Printing count as money in the United States" [@searle1995, 28]. X can be replaced by predicates that refer to any ontological category [@guala2015, 470].

According to RiE, institutions are norm-governed social practices. And Hindriks [-@hindriks2019] defines a social practice as a regularity in behavior that involves norms. Practices arise in response to signaling devices, which are salient features of the environment that enable agents to align their behaviors in beneficial ways, creating new strategies and thus giving rise to conventions—mutually beneficial behavioral regularities. Interdependent behavioral regularities in coordination games arise from signaling rules of the form "if D, do A", whereby agents condition their behavior on a signal to coordinate mutually beneficial interactions and achieve collective benefits. For example, a traffic light serves a signaling device that helps to make traffic safer and more efficient by coordinating behaviors.

Guala notes that social norms fulfill two functions highlighted by @north1990: they make behavior more stable and more predictable. However, as noted by Searle, they introduce new behaviors, as well, and they do it by changing game payoffs. Norms help explain not only the persistence of institutions, but also its emergence. It means that the 'scope of actionable signs' and normativity of social institutions as its core feature distinguishing it from 'animal conventions' might be connected.

On the game-theoretic account used in RiE, social norms are modeled as sanctions with costs that can alter behavior by influencing agents' preferences, as agents face costs for not conforming to it. High costs and a greater likelihood of violation-detection increase the incentive to cooperate. Institutions, in their turn, are maintained partly because of these norm costs.[5] Introduction of sufficiently high δ-parameter into cooperation problems transforms them into coordination ones. For instance, given a Prisoner's dilemma with a high δ-parameter representing a cost for a norm violation, the game becomes that of coordination with two equilibria — CC and DD instead of only DD [@crawford1995]. This shows that normative rules can be coordination devices, or "choreographers", as Gintis puts it [@gintis2009a]. How do the δ-parameter and normativity it represents emerge? And should the Grazing game have these parameters as well?

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

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

Table 5: Delta parameter transforming cooperation game into coordination game.

Before tackling these questions, let us first examine the backbone of 'scope of actionable signals' argument, the notion of bourgeois equilibrium.

3. Correlation and asymmetry of strategies

Guala and Hindriks draw inspiration for RiE in Maynard Smith's concept of *"bourgeois equilibrium"* [@maynard-smith1982]. They see the "Hawk-Dove-Bourgeois" (HDB) game of animal territorial ownership as representing a prototypical "animal convention". According to the authors, bourgeois equilibrium (BE) is essentially a CE, however Maynard Smith uses bourgeois to define evolutionary stable strategies ESS. It creates tension, for CE and ESS are mathematically distinct: the former is "too loose" and the latter is "too strict" in terms of the stability conditions, and it is unclear how they can be combined. Hence, the issue consists of clarifying the status of BE: whether it is a CE, an ESS or something else. It is due to being at the core of Guala's argument for institutions as correlation of strategies rooted in evolution. Let us look at the Maynard Smith's notion of BE captured in the the HDB game.

4.1 Maynard Smith's "Hawk-Dove-Bourgeois" game

@maynardsmith1982 famously has introduced the notion of ESS into game theory. A 'strategy' is a behavioral phenotype, a specification of what an individual will do in any situation. An ESS is a strategy that, if adopted

[5] However, as Hindriks [-@hindriks2019] argues, the costs are insufficient and there should also be normative beliefs.

by all members of a population, prevents the invasion of any mutant strategy by natural selection. The concept originated in the context of animal behavior, but can be applied to any phenotypic variation; e.g., growth form, age at first reproduction, or relative number of offspring

He proposes a model of a 'Hawk-Dove' game that represents a *mis*coordination game between two agents. In a competition for some resource, 'Hawk' fights for it and 'Dove' displays and retreats if threatened.

	Hawk	**Dove**
Hawk	$\frac{1}{2}(V-C)$, $\frac{(V-C)}{2}$	V
Dove	0	$\frac{V}{2}$

Table 6: A 'Hawk-Dove' game. The payoffs are determined by the value of the resource (V) and the cost of fighting (C). Value V increases the Darwinian fitness of an individual if they obtain the resource, and cost C reduces it if injured in a fight over the resource. Not gaining V, however, does not mean zero fitness.

As this model is at the core of Guala's theory, its assumptions are important. This model assumes an infinite population with asexual reproduction and symmetric contests between two opponents. It also has a finite set of strategies.

Defining the stability criteria for the strategies, he proposes that If a strategy I is stable against J, it must satisfy the "standard conditions" from @smith1973: the fitness of typical members adopting I must be greater than any mutant J, such that:

- either $E(I,I) > E(J,I)$ or $E(I,I) = E(J,I)$
- and $E(I,J) > E(J,J)$.

According to these conditions, D cannot be an ESS, for $E(D,D) < E(H,D)$, and H is an ESS if costs of injury are less than potential gain from the resource, $V > C$. If $V < C$, neither H nor D is an ESS. To proceed, Maynard Smith considers the behavior of individuals who can play either strategy with a certain probability, which they pass on to their offspring. This strategy takes the form 'play H with probability P, and D with probability $(1-P)$'.

A mixed strategy I, which randomly chooses an action from a set of possible actions, may be an ESS if the expected payoffs of the strategies composing it are equal. This follows from a theorem by @bishop1978: if a mixed ESS includes the pure strategies $A, B, C, ...$ with non-zero probability, then

$$E(A,I) = E(B,I) = E(C,I) = ... = E(I,I)$$

Intuitively, this means that if $E(A,I) > E(B,I)$, adopting A more often and B less often would be more advantageous than following strategy I, making it not an ESS.

However, I can be an ESS if probability of its adoption is $P = V/C$. In contests where the cost of injury is greater than the rewards of victory, $V < C$, mixed strategies with $P = V/C$ are evolutionarily stable.

What is important, a game with two pure strategies always has an ESS, and games with three or more strategies may not have one. As we remember, both "Hawk-Dove" and "Grazing" games have three strategies.

@maynardsmith1982 introduces the distinction between symmetric and asymmetric games. He illustrates them with animal contests. An asymmetric contest is one where participants have different roles, allowing them to use different strategies. Roles must be identifiable and can be based on gender, ownership, or intruder status. Circumstances which determine an individual's role are assumed to be independent of their genetic strategy. A contest with no role differentiation is 'symmetric'. @maynardsmith1982 characterizes them as follows:

1. Contests are between two individuals of distinct roles (e.g., owner/intruder, larger/smaller, older/younger);
2. Both individuals know their role;
3. Both have the same strategies available (e.g., escalate, retaliate, display);
4. Role may influence chances of winning or value of victory.

The Hawk-Dove game is symmetrical—both players have the same choice of strategies and payoffs. However, most contests are asymmetric, with differences in size, strength, gender, age, or ownership influencing strategy choice and/or altering payoffs or success in escalation. Even when the asymmetry does not change payoffs or escalation outcomes, it may still determine the players' actions.

In this example, the Hawk-Dove game is extended to include a third strategy, B (or Bourgeois), which is defined as 'if owner, play Hawk; if intruder, play Dove'. This strategy is ESS and the only ESS of this game. It is assumed that each strategy type is owner and intruder equally frequently. Hence, even when ownership does

	Hawk	**Dove**	**Bourgeois**
Hawk	-1	2	0.5
Dove	0	1	0.5
Bourgeois	-0.5	1.5	1

Table 7: 'Hawk-Dove-Bourgeois' game

not alter payoffs or success in fighting, an asymmetry of ownership can be used as a conventional one to settle the contest.

Here, the B player chooses H and D with equal frequency, acting as an owner on half the occasions and an intruder on the other half. And when two B's compete, if one chooses H, the other chooses D. If $V > C$, the ESS is H as it is worth risking injury to gain the resource; if $V < C$, the ESS is B as ownership settles the contest without escalation. It means that in both 'Hawk-Dove' and 'Grazing' games $V < C$.[6]

Crucially, this assumes that *the probability of an individual occupying a role is independent of their strategy.* This holds true even for strategy B, wherein the individual's role is correlated with their chosen action (Hawk or Dove). The assumption is that the strategy B itself is unrelated to role. In other words, If an agent is indeed an 'owner', it does not entail that she always plays a certain 'owner' strategy like 'Hawk' or 'Bourgeois'. However, according to @gintis2007a, empirical findings corroborate the existence of the 'endowment' effect, when owners value a resource more than intruders, thus making them fight harder for it. It presupposes a certain degree of correlation between role and strategy.

@smith1976 used the term 'uncorrelated asymmetry' to refer to payoff-irrelevant differences in an otherwise symmetric game. In asymmetric contests the value of the resource, or chance of victory, is not the same for both owner and intruder. The payoffs to owners and intruders are often not equal, so the territory may be more valuable to an owner who has already familiarized themselves with food, refuge, and other. Ownership may even offer advantages in escalated contests. Inequality of payoffs is possible due to size or age asymmetry. Even if there is no inequality, an asymmetry can still settle contests. Thus, "Grazing" game as presented by Guala, does not require a correlated device and may be solved by uncorrelated asymmetry alone, as both players recognize the asymmetry of ownership and the value of territorial gains is less than the costs of potential injury, $V < C$, for they might value their own territory more than potential one.

It is interesting that @maynardsmith1982 considers the 'social contract' game as one which humans can play but animals cannot. This game involves a group of individuals agreeing on a behavioral regularity and punishing any member who deviates from it. However, the act of punishing carries a cost, so in order to maintain stability, refusal to participate in enforcement must be considered a breach and punished as well. To ensure enforcement, a subgroup may be rewarded for carrying it out. Essentially, the 'social contract' game differs from any other coordination or cooperation game in added normativity, that can be represented by a delta parameter. And this normativity is characteristic of social institutions as defined in rules-in-equilibria theory. This renders Guala's example with 'Grazing game' problematic, for, following Maynard Smith, not only HDB and 'Grazing game' are conceptually distinct, for they presuppose different cognitive capacities, but they must have different payoff structure due to added normativity in the human case.

Overall, BE assumes that each player is trying to maximize their own self-interest, but no player is attempting to dominate or exploit the others. A BE is certainly a situation and not a solution concept. It occurs when the players have reached a strategy profile in which none of them can improve their payoff by changing only their own strategy, while also recognizing the other player strategies.

As there are two possible interpretation of BE—correlation of strategies and uncorrelated asymmetry, let us consider both.

4.2 Interpretation of HDB: correlation of strategies

Guala and Hindriks put forward that coordination in social institutions, and in 'Hawk-Dove-Bourgeois' as an exemplar case of property, is due to correlation of strategies. But what is "correlation of strategies" in the first place?

Correlation of strategies is a stable state of strategic interaction. It is represented by the concept of correlated equilibrium (CE) that goes beyond the Nash equilibrium and allows players to coordinate their strategies through

[6]However, it is still not clear whether human players such as grazers have genuine fitness rather than utility value function. As @sterelny2012 suggests, there has been an evolutionary shift from fitness to utility correlated with the demographic explosion in the Pleistocene and subsequent significant decline in individual-level heritability of cultural traits, for offspring did not more resemble their parents informationally and ideologically due to the abundance of cultural information sources.

the use of a common randomizer, such as a coin toss or a dice roll. This allows players to make decisions based on their beliefs about how the other players will act, which can increase the efficiency of their strategies. The concept of CE has been used to explain various phenomena in strategic decision making, including how people form coalitions, how firms cooperate and compete, and how players interact in team games.

Formally, a correlated equilibrium is a probability distribution p over the set of action vectors S if the strategy vector τ^* is a Nash equilibrium of the game $\Gamma^*(p)$ [@zamir2013, 307]. In other words, for every player $i \in N$:

$$\sum_{s_{-i} \in S_{-i}} p\left(s_i, s_{-i}\right) u_i\left(s_i, s_{-i}\right) \geq \sum_{s_{-i} \in S_{-i}} p\left(s_i, s_{-i}\right) u_i\left(s_i', s_{-i}\right), \quad \forall s_i, s_i' \in S_i$$

The equation states that the optimal strategy for each player is dependent on both their own decisions and on those of other players, which reflects how CE allows players to take into account each other's behavior when making decisions.

The key feature and difference of CE is randomization. As @aumann1987 points out, correlation is more general than mixing of strategies, for the latter can be formally seen as the former by considering the product probability space $\Gamma^1 \times \cdots \times \Gamma^n$, where Γ^i is the set of outcomes corresponding to player i's mixed strategy. Players make correlated, or nonindependent, choices when they observe the same random variable.

Back to the HDB game, its important feature is that if $V < C$, the B strategy helps to settle contests *conventionally*. Maynard Smith does not emphasize the notion of convention, but it is key in Guala's discussion of social institutions as it describes mutually beneficial behavioral regularities. What Maynard Smith means by conventional settlement is that there is shared 'understanding' of the situation between the players that helps to decide on the action. But what precisely does 'conventional settlement' mean regarding the B strategy in the HDB game? Let us start with convention as CE.

@vanderschraaf1995 formalizes Lewis's notion of salience in coordination games and models conventions as correlated equilibria instead of Nash ones.

@lewis2008, building on the ideas of @schelling1980, proposed the notion of salience as an explanation of how a convention become established. A coordination equilibrium[7] of a game is salient if it is noticeable to all players, and they expect their opponents to choose the same equilibrium, resulting in them playing it. As Lewis suggests, salience can be determined by environmental factors.

Lewis considers a coordination equilibrium a convention if the players have common knowledge of a mutual expectations criterion (MEC). It means that each agent has a decisive reason to conform to her part of the convention, expecting the other agents to do likewise. He states that an equilibrium must be a coordination equilibrium to reflect the notion that a person conforming to a convention wants their intention to be seen as such. Vanderschraaf calls it the public intentions criterion (PIC). Furthermore, Lewis argues that common knowledge of the MEC is necessary for a convention. However, as Vanderschraaf notes, it is not sufficient, since common knowledge of the MEC can be satisfied at any strict Nash equilibrium.

Vanderschraaf defines a convention as a mapping of "states of the world" to strategy combinations of a noncooperative game [@vanderschraaf1995, 69]:

DEFINITION 1. A *game* Γ is an ordered triple $(N, S, \mathbf{u})$ consisting of the following elements:

1. A finite set $N = \{1, 2, ..., n\}$, called the *set of players*;
2. For each player $k \in N$, there is a finite set $S_k = \{A_{k_1}, A_{k_2}, \ldots, A_{kn_k}\}$, called the *alternative pure strategies* for player k. The Cartesian product $S = S_1 \times \cdots \times S_n$ is called the *pure strategy set* for the game Γ;
3. A map $\mathbf{u} : S \to \mathbb{R}^n$, called the *payoff function* on the pure strategy set. At each strategy combination $\mathbf{A} = (A_{1j_1}, \ldots, A_{nj_n}) \in S$, player k's payoff is given by the kth component of the value of $\mathbf{u}$, that is, player k's payoff u_k, at $\mathbf{A}$ is determined by

$$u_k(\mathbf{A}) = I_k \circ \mathbf{u}(A_{1j_1}, \ldots, A_{nj_n}),$$

where $I_k(\mathbf{x})$ projects $\mathbf{x} \in \mathbb{R}^n$ onto its kth component.

[7] Coordination equilibrium is a concept defined by philosopher David Lewis which states that when two or more individuals are engaged in a coordination game, they will naturally gravitate towards the same outcome, as this is the most rational choice. The idea is that each individual will tend to choose the same outcome because they can both benefit from it. This is in contrast with Nash equilibrium, where each individual must make a choice that maximizes their own payoff without considering the other's payoff.

As Vanderschraaf builds on Aumann's model [-@aumann1987], each player has a personal *information partition* $\mathcal{H}_k$ of a probability space Ω. Elementary events on Ω are called *states of the world*. At each state ω, each player k knows which element $H_{kj} \in \mathcal{H}_k$ has occurred, but not which ω. H_kj represents k's private information about the states of the world. While k knows the opponent partitions, she does not know their content. A function $f : \Omega \to S$ defines a *exogenously correlated strategy n-tuple*, such that at each state of the world $\omega \in \Omega$, each player k selects a strategy combination $f(\omega) = (f_1(\omega), ..., f_n(\omega)) \in S$ correlated with the state of the world ω. Thus, by playing $f_k(\omega)$, k follows *Bayesian rationality* and maximizes expected payoff given private information and expectations regarding opponents.

DEFINITION 2. Given $\Gamma = (N, S, \mathbf{u})$, Ω, and the information partitions $\mathcal{H}$ of Ω as defined above, $f : \Omega \to S$ is a *correlated equilibrium* if and only if, for each $k \in N$,

1. f_k is an $\mathcal{H}_k$-measurable function, that is, for each $H_{kj} \in \mathcal{H}_k$, $f_k(\omega)$ is constant for each $\omega' \in H_kj$, and
2. For each $\omega \in \Omega$,
$$E(u_k \circ f|\mathcal{H}_k)(\omega) \geq E(u_k \circ (f_{-k}, g_k)|\mathcal{H}_k)(\omega)$$

where E denotes expectation, '$-k$' refer to the result of excluding the kth component from an n-tuple. This holds for any $\mathcal{H}_k$-measurable function $g_k : \Omega \to S_k$. The correlated equilibrium f is *strict* if and only if the inequalities are all strict.

The measurability restriction on f_k means that k knows her strategy in each ω. This definition implies that players have common knowledge of the payoff structure, partitions of Ω, and $f : \Omega \to S$, which is needed to compute expected payoffs and reach correlated equilibrium. In addition, if the players possess common knowledge of Bayesian rationality, they will follow their ends of f, expecting others to do the same, since they jointly maximize expected utility in this way.

The agents refer to a common information partition of the states of the world. While each agent k has a private information partition $\mathcal{H}_k$ of Ω, there is a partition of Ω, namely the intersection $\mathcal{H} = \cap_{k \in N} \mathcal{H}k$, of the states of the world such that for each $\omega \in \Omega$, all the agents will know which cell $H(\omega) \in \mathcal{H}$ occurs. The agents' expected utilities in the following Definition 3 are conditional on their common partition $\mathcal{H}$, reflecting the intuition that conventions rely upon information that is public to all.

The agents' expected utilities are conditioned on their common information common partition $\mathcal{H}$ of the states of the world, which is the intersection of all their private partitions $\mathcal{H} = \cap_{k \in N} \mathcal{H}_k$. This reflects that conventions depend on information available to all agents.

DEFINITION 3. Given $\Gamma = (N, S, \mathbf{u}), \Omega$, and the partition $\mathcal{H}$ of Ω of events that are common knowledge among the players, a function $f : \Omega \to S$ is a convention if and only if for each $\omega \in \Omega$, and for each $k \in N$, f_k is $\mathcal{H}$-measurable and

$$E\left(u_k \circ f \mid \mathcal{H}\right)(\omega) > E\left(u_k \circ \left(f_{-j}, g_j\right) \mid \mathcal{H}\right)(\omega)$$

for each $j \in N$ and for any $\mathcal{H}$-measurable function $g_j : \Omega \to S_j$.

It means that if any player j deviates from a convention f, every player $k \in N$, including j, will be worse off. This definition of convention as a strict correlated equilibrium satisfies the PIC, as all agents are aware of the common partition and the strategies each player is expected to play. Thus, if any opponent mistakenly thinks that a player k will play a strategy $g_k(\omega) \neq f_k(\omega)$ other than the one prescribed by f, they may be tempted to deviate, resulting in a worse-off outcome for k. Conversely, if all opponents are aware that k will play her strategy $f_k(\omega)$ at each state of the world $\omega \in \Omega$, then they have a strong incentive to conform with convention $f(\omega)$, which gives k an improved outcome.

Overall, Vanderschraaf's contribution is formalization of salience, hence he uses the *common* information partition $\mathcal{H}$ as a necessary restriction to make the definition of convention conform with Lewis' spirit. The other question is how salience itself emerges. Lewis suggests that pre-game communication, precedent, and environmental cues may lead agents to link their expectations and actions with various "states of the world", thus achieving correlated equilibrium. However, these sources of salience face the problem of infinite regress, for it is unclear how precedent or pre-game communication occurred in the first place without an established and shared conventional rules. Vanderschraaf, along with Skyrms [@vanderschraaf1993], proposes *inductive deliberation* as a mechanism by which salience is being established. It requires agents to be Bayesian rational and works by recursive belief modification. Players can reach a correlated equilibrium without communication by dynamically updating their beliefs using a common inductive rule, even if their beliefs don't initially allow for an equilibrium.

What is important in regard to the B strategy in the HDB game, Vanderchraaf notes that conventions as correlated equilibria allow for characterization of a wide range of equilibria. Given a game Γ with pure strategy coordination equilibria $\mathbf{A}_1, \ldots, \mathbf{A}_m, m \geqslant 2$, and a lottery Ω with mutually exclusive outcomes $H_1, \ldots, H_m$ such that $p_k\left(H_j = \lambda_j\right)$ for each player j. Then if the players condition on $\mathcal{H} = \{H_1, \ldots, H_m\}$, and $f : \Omega \to S$ is defined by $f(\omega) = \mathbf{A}_j$ if $\omega \in H_j$, then [[Convention is CE, as salience is an information partition#^6afd45|inequality]] is satisfied for all $\omega \in \Omega$, making f a convention. With infinitely many possible values for the λ_j's, any noncooperative game with two or more pure strategy coordination equilibria has infinitely many correlated equilibria corresponding to conventions.

Convention as correlated equilibrium allows for the "fair" coordination, even though no pure strategy equilibrium exists. Consider the "Battle of Sexes" game.

	$A1$	$A2$
$A1$	$10, 7$	$0, 0$
$A2$	$0, 0$	$7, 10$

Table 8: "Battle of sexes" game

Neither of the pure strategy Nash equilibria in this game is "fair", in the sense that the players receive the same payoff. This game has a mixed Nash equilibrium at which Player 1 plays $A1$ with probability $\frac{2}{3}$ and Player 2 plays $A2$ with probability $\frac{2}{3}$, and at this equilibrium each player's expected payoff is $\frac{2}{3}$, so this equilibrium is "fair". However, at the mixed Nash equilibrium, both players are indifferent to the strategies they play given what each player believes about her opponent, so this equilibrium fails the PIC and is consequently not a convention. Nevertheless, there is a correlated equilibrium fair to both players, and which each player will prefer over the pure strategy equilibrium that is unfair to her.

This game has a mixed Nash equilibrium at which both agents play their strategies with probability $\frac{2}{3}$, yielding an expected payoff of $\frac{2}{3}$ for each agent. However, this equilibrium does not satisfy the PIC and is thus not a convention. Nevertheless, there is a correlated equilibrium that is fair to both players and preferable to the pure strategy equilibrium.

With a toss of a fair coin, there is a probability space $\Omega = \{H, T\}$ with "heads" and "tails". The agents have a common information partition $\mathcal{H} = \{\{H\}, \{T\}\}$ and the correlated strategy combination is denoted as a function $f : \Omega \to \{A1, A2\} \times \{A1, A2\}$ with $f(H) = (A1, A1)$ and $f(T) = (A2, A2)$. Player 1 has a higher expected payoff with this combination than any of the other strategies, so she will not deviate from it. The expected payoff for Player 1 is 2 if the outcome is H, and 1 if it is T.

$$
\begin{aligned}
&E\left(u_1 \circ f \mid H\right) = 2 > 0 = E\left(u_1(A2, A1)\right) \mid H\right), \text{ and} \\
&E\left(u_1 \circ f \mid T\right) = 1 > 0 = E\left(u_1(A1, A2) \mid T\right)
\end{aligned}
$$

The same holds for the second player. To this end, neither player would want to deviate, since the overall expected payoff at this equilibrium for each player is

$$
E\left(u_k \circ f\right) = \frac{1}{2} \cdot E\left(u_k \circ f \mid H\right) + \frac{1}{2} \cdot E\left(u_k \circ f \mid T\right) = \frac{3}{2}
$$

It means that each player prefers the expected payoff from f to that of the mixed equilibrium.

One intrinsic problem with BB as CE, however, is the source of randomization. Some scholars appeal to Nature as to a such source, calling it a *correlation device*, thus eliminating the tension between the requirement of randomization and symmetry of ESS [@cripps1991; @skyrms2014; @metzger2018]. In particular, Gintis defines CE as an NE of a game G augmented by the *initial move by Nature* that who observes a random variable γ on a probability space (Γ, p) and issues directives $f_i(\gamma) \in S$ to each player i, such that choosing the directive is the best response given agents having a common prior p and assuming other players are also following Nature's directives [@gintis2009a, 135-136]. The crucial assumption, though, is that the game is epistemic and has common priors. This implies that all agents agree on the probability distributions over their actions and have *joint randomized probability*. Assuming a strict correlated equilibrium—that each agent i has a single best response $s_i(\omega)$ in every state ω—each player's move is known to the others and rationality dictates that each must play a best response to the actions of the others.

In their theory, Guala and Hindriks endorse this interpretation of the "Hawk-Dove".

According to Skyrms, the implicative nature of the B strategy is genuinely correlative. According to Skyrms, the B, B strategy profile is CE 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" [-@skyrms2014, 78].

However, ... [[Symmetry broken by chance events leads to correlated equilibrium in iterated Hawk-Dove game]]

Maynard Smith's notion of a "bourgeois" equilibrium and the concept of correlated equilibrium are related but not identical.

In game theory, a Nash equilibrium is a set of strategies where no player can improve their outcome by unilaterally changing their strategy. A correlated equilibrium is a probability distribution over joint strategies that has the property that each player's strategy is optimal given their information about the distribution.

In contrast, Maynard Smith's idea of a "bourgeois" equilibrium refers to a situation where all players play the same mixed strategy, such as choosing actions uniformly at random. In this case, no player has an incentive to deviate from their strategy, since any deviation would lead to a lower payoff.

However, it is possible for a "bourgeois" equilibrium to also be a correlated equilibrium. For example, suppose there are two players in a game, A and B, and they both choose their actions uniformly at random. This leads to a certain probability distribution over joint strategies, which could be a correlated equilibrium if no player has an incentive to deviate from their strategy given their information about the distribution.

So while Maynard Smith's notion of a "bourgeois" equilibrium is not identical to the concept of correlated equilibrium, the two ideas can overlap in certain situations.

Maynard Smith's notion of a "bourgeois" equilibrium is not always identical to correlated equilibrium, and there are situations where the two concepts differ.

One situation where the two ideas can overlap is when all players have the same best response to each possible correlation. In this case, any correlated equilibrium must also be a "bourgeois" equilibrium. For example, consider the following coordination game:

Player 1 chooses between A and B Player 2 chooses between X and Y

The payoffs are as follows:

(A,X): (5,5) (A,Y): (0,0) (B,X): (0,0) (B,Y): (3,3)

In this game, both players prefer to coordinate on (A,X) since it yields the highest payoff for both. If both players choose their actions uniformly at random, the resulting probability distribution is a correlated equilibrium where each player has an equal chance of playing A or B, and X or Y. This distribution is also a "bourgeois" equilibrium since all players play the same mixed strategy.

However, there are situations where the two ideas diverge. One such situation is when different players have different best responses to each possible correlation. In this case, there may be correlated equilibria that are not "bourgeois" equilibria, or "bourgeois" equilibria that are not correlated equilibria. For example, consider the following sequential game:

Player 1 chooses between A and B Player 2 observes Player 1's choice and then chooses between X and Y

The payoffs are as follows:

(A,X): (10,5) (A,Y): (0,0) (B,X): (0,0) (B,Y): (5,10)

In this game, if Player 1 chooses A, then Player 2 should choose X to maximize their payoff, while if Player 1 chooses B, then Player 2 should choose Y. There is no equilibrium where all players play the same mixed strategy, so there is no "bourgeois" equilibrium in this game. However, there is a correlated equilibrium where Player 1 chooses A with probability 2/3 and B with probability 1/3, and Player 2 chooses X with probability 2/3 and Y with probability 1/3. In this equilibrium, each player's strategy is optimal given their information about the distribution, but the resulting correlation is not a "bourgeois" equilibrium since the players' mixed strategies are different.

In game theory, the correlation of strategies and uncorrelated asymmetry are related in the sense that they can affect the outcome of a game and the strategies that players choose to play.

Correlation of strategies refers to situations where players coordinate their actions before the game starts by sharing information or using other signaling mechanisms. This allows them to reach an agreement about which strategy to adopt based on shared expectations of what other players will do. Correlated equilibria are solution

concepts that allow for such coordination. In a correlated equilibrium, players use the correlation device to select a strategy based on the signal they receive, and no player has an incentive to deviate from this strategy.

Uncorrelated asymmetry, on the other hand, refers to differences in players' preferences or abilities that are not observable to the other players. These differences can arise due to factors such as differences in initial endowments, skill levels, or beliefs. When there is uncorrelated asymmetry in a game, each player has private information that affects their choice of strategies. This can make it difficult for players to coordinate their actions and can lead to inefficiencies in the outcome of the game.

However, if the uncorrelated asymmetry is made public through some correlation device, such as a signal or a mediator, then players can use this information to coordinate their actions more effectively. For example, in an auction where bidders have different valuations for the item being sold, a common correlation device is to reveal the highest bid to all bidders. This allows bidders to revise their bids based on the new information and can lead to a more efficient outcome.

Overall, the relationship between correlation of strategies and uncorrelated asymmetry is complex and can depend on the specific context of the game being played. However, in general, correlation devices can help overcome uncorrelated asymmetries and facilitate better coordination among players. ***

4.3 Interpretation of HDB: uncorrelated asymmetry

Another interpretation of HDB involves uncorrelated asymmetry instead of correlation. @oconnor2019 employs this interpretation in her treatment of emergence of unfairness due to social categories as solutions to inherently institutional coordination problems. On this account, HDB strategy profiles are based not on correlation, but on uncorrelated asymmetry. It is a feature of games where players extract additional information from environment not included in the structure of a game. For example, they know that they are "Hawks" or "Doves" rather than their strategies are randomized. This underlies an important methodological distinction between correlated equilibrium and evolutionary stability.

5. Evolution, Bayesian updating and correlation

Conclusion

References