Are Institutions Rules in Equilibrium? Comments on Guala’s Understanding Institutions

Wlodek Rabinowicz

Abstract
In this comment on Francesco Guala’s Understanding Institutions, I express my admiration for the book but I also raise some critical criticisms: His general account of institutions as rules-in-equilibrium seems to get their ontology wrong by disregarding their material side - their concrete realizations. It also disregards social institutions whose rules are not (and are not meant to be) in equilibrium. Finally, his suggestion that institutional equilibria necessarily involve correlation devices appers to lack justification.

Keywords
Guala, institutions, social institution, rules, equilibria, correlated equilibrium, Searle, Lewis, Aumann

Francesco Guala has written an excellent book, both thought provoking and entertaining. He writes in a clear, accessible way; puts forward exciting general claims; and illustrates them with striking examples. The book should appeal to

Received 10 August 2018

1Lund University, Lund, Sweden

Corresponding Author:
Wlodek Rabinowicz, Department of Philosophy, Lund University, 221 00 Lund, Box 192, Sweden.
Email: wlodek.rabinowicz@fil.lu.se
a broad academic audience. Philosophers, economists, sociologists, political scientists, and psychologists should find it rewarding, even if they are not conversant with the area. This is certainly true in my own case. The topic Guala writes about—the nature of social institutions—is not a subject I am closely familiar with. I have not followed the literature in this area nor given sustained thought to the problems that arise. For this reason, I feel somewhat apprehensive writing these comments; the criticisms I have might be rooted in misunderstandings on my part. If so, I hope my worries will be laid to rest. The questions I would like to discuss concern two main issues: (a) Guala’s general account of the nature of institutions and (b) some game-theoretic aspects of this account. As for (a), I want to question Guala’s definition of institutions, which seems to get their ontology wrong: by focusing on structural aspects, it ignores the “materiality” of institutions—the fact that they are objects involving both people and other material components. As for (b), I want to examine Guala’s suggestion that coordination problems require correlation devices for their successful solution and that institutions therefore necessarily involve so-called correlated equilibria.

1. The Nature of Institutions

There are two influential traditions that form the background of Guala’s proposal. On one hand, we have people like John Searle who interpret institutions as systems of rules (cf. Searle’s *The Construction of Social Reality* 1995), on the other the tradition that takes its departure from David Lewis who interprets institutions as game equilibria (cf. Lewis’s *Convention* 1969). Guala’s goal is to combine these two traditions. He defines institutions as “rules that people are motivated to follow,” or—as he very often puts it—as “rules-in-equilibrium”:

> By combining the rules account with the equilibria account we obtain a unified theory that I call the *rules-in-equilibrium* approach to the study of institutions. Rules by themselves lack the power to influence behavior, but together with the right system of incentives and beliefs, they can influence the behavior of large groups of individuals. Institutions, in a nutshell, are rules that people are motivated to follow. (XXV)

This unified theory combines the advantages of the rules account with those of the equilibrium account.² The rules account, Guala suggests, squares well with our ordinary way of thinking about institutions:

²In Guala’s account, the rules that define an institution regulate behavior of the participants. Rules that Searle focuses on do not seem to have this character. They typically are of the form: “X counts as Y in circumstances C” (e.g., “Pieces of paper of this kind count as money”). In chapter 5, however, which is based on joint work with Frank Hindriks, Guala argues that such “constitutive” rules are reducible to ordinary rules of behavior.
The rules account is close to our vernacular, prescientific understanding of institutions: intuitively, institutions regulate behavior, making certain actions appropriate or even mandatory in specified circumstances. The institution of private property, for example, regulates the use of resources by indicating who has access to them. The institution of money regulates the use of paper certificates in economic transactions. And the institution of marriage regulates the behavior of two or more individuals who pool their resources to raise kids, manage property, and help each other in many different ways. (XXIV)

The equilibrium account, on the other hand, focuses on what makes us follow the rules and thereby makes institutions exist as parts of social reality:

But if institutions are rules, how do they influence behavior? Stating a rule is clearly insufficient to bring about an institution. To realize why, consider that there are plenty of ineffective rules: rules that are officially or formally in existence but that are nevertheless ignored . . . There must be some special ingredient that makes people follow the rules . . . .

The equilibria account of institutions tells us what the special ingredient is: effective institutions are backed up by a system of incentives and expectations that motivate people to follow the rules. An equilibrium in game theory is a profile of actions or strategies, one for each individual participating in a strategic interaction . . . each strategy must be a best response to the actions of the other players or, in other words, . . . no player can do better by changing her strategy unilaterally. If the others do their part in the equilibrium, no player has an incentive to deviate.

Since the actions of a strategic game can be formulated as rules, equilibrium-based and rules-based accounts of institutions are compatible. (XXIV-V)

The equilibrium concept appealed to in this quote is what game theorists call a “Nash equilibrium.” As we will see in the following section, Guala thinks that another equilibrium concept is even more apposite in the analysis of institutions. But this matter can wait for now.

The rules-in-equilibrium account of institutions seems quite attractive, but at the same time somewhat strange. In some respects, it appears to be too strong, to require too much, while in other respects it seems too thin.

Let us consider two examples of institutions.

1.1. Example 1: A Society’s Morality

This social institution can, I think, be understood as a set of rules (even though ethical particularists might question this claim). So far, so good, then. But are they rules that people are motivated to follow? Are they not rather rules that people believe they ought to follow? Typically, the rules we
effectively are motivated to follow are considerably less stringent than what morality requires. Thus, for example, in the Prisoner’s Dilemma type of situations (PD-games), moral rules prescribe cooperation. This means they are not rules-in-equilibrium: cooperation is not a Nash equilibrium solution of a PD-game. The equilibrium rule prescribes defection to each player. This is also what we typically are motivated to do in such a game, at least as long as we do not expect the game to be repeated and thus need not worry about reputation effects.

Does this show that the societal morality is not an effective, well-functioning social institution? I do not think so and I do not think Guala would want to make such a claim. Indeed, it is arguable that the morality we accept and share does influence and shape our behavior in many important ways. It is only that this influence is not a direct and straightforward one. While the moral rules we accept do modify our behavior, only few of us fully live up to our moral commitments. Many moral rules are “aspirational” in nature: they express ideals that, if we are acculturated enough, we try to approximate. But we regularly fall short of these ideals in our actual behavior. And if we all did act as morality requires, the resulting state would not form an equilibrium.3 The equilibrium state is one in which all of us accept the moral rules, but not the one in which we all follow the rules in question.

Similar remarks apply, to some extent at least, to such social institutions as legislation, religion, or etiquette.

1.2. Example 2: A University

The University of Milan, at which Guala works, is one of his own examples of an institution. Indeed, it is the example with which he starts the book:

... most of the things that we see are institutional entities. An “institutional entity” is an object with properties or characteristics that depend on the existence of an institution. Antonio, for example, is a colleague of mine because we are both employees of the same university, and the University of Milan is an institution. (XVII)

But how does the University of Milan fit Guala’s rules-in-equilibrium account? Certainly, no one would be willing to equate this institution with a set of rules that (some) people are motivated to follow! A university is so much more than

3Or, even if it would, as might be the case with some moral rules, the resulting equilibrium would not necessarily be evolutionarily stable. A population of perfectly moral agents might be vulnerable to invasions of free-riders. I disregard in what follows this evolutionary perspective, since Guala does not adopt it in his book.
that. While it is regulated by rules, both formal and informal, it consists of individuals (students, professors, administrative staff), buildings, labs, and so on. Given the university’s existence, these individuals and things have the status of institutional entities. As such, they are parts of the university.

Members of the university are motivated to follow certain rules of behavior in their academic activities, but these rules at best form the way in which a university is organized. They regulate its members’ behaviors, but they certainly are not the university itself.

So what could Guala mean? Judging from his discussion in the book, what he might mean is that an institution such as a university is a game of some sort (or, better, a repeated game), with

- a number of players (students, professors, administrative staff, etc.),
- a set of strategies (rules) for each player to choose from,
- a matrix of payoffs for the players, which depend on the strategies they all choose,
- a certain specific equilibrium solution of this game.

Or, at least, this is so for stable institutions. In an unstable institution, the chosen strategies (rules) are not in equilibrium, I suppose.

Also, the important elements of the game are

- mutual beliefs of the players—beliefs about the other players and their expected choices.

Such beliefs support and justify the particular equilibrium solution of the game in question.

Now, this game has a material base, which includes both persons and things. Its players are not abstract individuals, mere institutional entities, but concrete persons, and the game is played out in a concrete environment, in seminar rooms, lecture halls, university restaurants, offices, and labs. This material base and its components—people and things—should also be counted as parts of the institution. This, at least, is clearly indicated by our ordinary way of talking and thinking about such institutions as universities.

But, one might object, do the components of the material base really belong to the institution? Cannot each of them be replaced or removed without the institution itself ceasing to exist? It is true, of course, but irrelevant. For an object to be a part of a whole it is not required for it to be an essential, irreplaceable part. Indeed, it is an essential feature of such institutions as universities that they contain inessential components. An institution cannot exist without such components, but none of them is necessary to its existence;
each of them is replaceable or can even be simply removed without replacement, without the institution ceasing to exist.

Thus, if I am right about all this, the identification of an institution with rules-in-equilibrium is, at best, misleading. It is a case of a *pars pro toto*. This kind of account focuses on just one, admittedly important, aspect of an institution, but not on the institution as a whole.4

Indeed, the nature of an institution such as a university seems even more complex than my earlier sketch would suggest. A university is not describable as one game, not even as one repeated game, but rather as a collection of many different games—which might have different formal structures and which might involve different groups of players. Some of these games are played by the members of an institution, while others involve both members and nonmembers: think of the interactions between academics and the society at large, or of the academics’ interactions with members of other universities.

Also, there is a need to bring in the aspect of change in connection with institutions: games played by the members of a university change over time. Or, even if they stay the same, the players might move to other equilibrium solutions. Just as the components in the material base are replaceable without the institution ceasing to exist, so are rules followed by the players.

Thus, the nature of an institution is highly complex. This complexity is kept hidden by such shorthand-characterizations as “rules-in-equilibrium.”

Similar remarks apply of course to many other institutions: prisons, hospitals, courts, and so on. They also apply to larger institutional wholes, such as, say, a country’s system of health care, its system of higher education, and so on.

At this point, however, one should note an important ambiguity in our usage of the word “institution.” A society’s morality, which was my first example, is a social institution in a different sense than a university or a prison. The former can be identified with a set of rules (even though these rules, as we have seen, need not be in equilibrium), but such an identification is out of place in the case of universities or prisons: it gets the basic ontology of the latter institutions wrong. The ambiguity in question is highlighted in the Oxford English Dictionary, which distinguishes between the following two senses of “institution” (along with several other senses that are not of interest here):

4An anonymous referee agrees with this criticism but notes that “it is the kind of criticism that can be levied at most, if not all, formal models.” It might well be so, but I think it is important that formal constructions, however idealized and simplified, avoid misrepresenting the basic ontology of the phenomena they model.
6a. An established law, custom, usage, practice, organization, or other element in the political or social life of a people; a regulative principle or convention subservient to the needs of an organized community or the general ends of civilization.

7a. An establishment, organization, or association, instituted for the promotion of some object, esp. one of public or general utility, religious, charitable, educational, etc., e.g. a church, school, college, hospital, asylum, reformatory, mission, or the like . . . .

A society’s morality exemplifies (6a), while a university is an example of (7a). These two senses of “institution” would require different ontological accounts. The rules-in-equilibrium account fits (6a) considerably better than (7a). An institution in the former sense does seem to be a set of rules, though not necessarily in equilibrium, but for the institutions in the latter sense, such characterization is grossly inadequate, as we have seen. However, Guala does not explicitly distinguish between these two senses of “institution” in his book.

**Correlated Equilibria**

The interactions that characterize an institution are (typically) coordination games, that is, games in which there are several Nash equilibria. This confronts the players with the problem of coordinating their strategies.

In Chapter 4 (“Correlation”), Guala exemplifies this problem with a version of the Chicken Game (or the Hawk-Dove Game, as it sometimes is called):

Both players in this game, Row and Column, face a choice between two strategies, \( D \) and \( C \), with \( D \) standing for *Dare* and \( C \) for *Chicken out*. One can

---

5OED online, accessed 23.02.18.

6Guala does distinguish between institutions as abstract *types* and as concrete *tokens*, that is, as particular instantiations of types (XXIf; see also chapter 14), but it is not the same distinction as the one I have in mind. Both my examples concern tokens: a particular university and the morality of a particular society. But the corresponding distinction could just as well be made at the level of types. A type of institution such as university or prison is very different from such types as societal morality or, say, legislation.

7As already mentioned, a Nash equilibrium is a profile of strategies, with one strategy for each player, such that by doing her part in the profile, each player maximizes her own payoff on the assumption that other players are doing their parts.
also read C as Cooperate and D as Defect.\textsuperscript{8} The payoffs to Row and to Column from each strategy combination are specified in the cells of Table 1. Thus, for example, the payoff from Row choosing D and Column choosing C is 2 for Row and 1 for Column. For each player, her payoffs are measured on an interval scale: the choice of the unit and the zero point are arbitrary, but the ratios of the differences between the payoffs are not. It is not necessary to assume interpersonal comparability, that is, that the scale on which the payoffs are measured is the same for Row and for Column, but we shall make this assumption in what follows.

D and C are the “pure,” basic strategies in this game, but the game also allows for “mixed” strategies, which are lotteries on pure strategies. Instead of choosing C or D, a player might opt for a lottery \((pD, (1 - p)C)\), which yields D with probability \(p\) and C with probability \(1 - p\). There are two Nash equilibria in pure strategies in this game, \((D, C)\) and \((C, D)\), which both give to the player who dares a higher payoff than to the one who chickens out (2 as compared to 1), and one Nash equilibrium in mixed strategies. The latter consists in each player opting for a lottery \(\left(\frac{1}{2}D, \frac{1}{2}C\right)\). The expected payoffs in this “mixed” equilibrium are the same for each player:

\[
\frac{1}{2}(\frac{1}{2}(0) + \frac{1}{2}(2)) + \frac{1}{2}(\frac{1}{2}(1) + \frac{1}{2}(1)) = 1.
\]

In view of this multiplicity of Nash equilibria, the players face the problem of coordination: it is not given what strategy Row should play to maximize her expected payoff, as long as it is not given what strategy Column is going to play. But Column’s strategy choice depends on her beliefs about Row’s choice, which in turn depends on Row’s beliefs about Column’s choice, and so on. The problem is how to break out of this vicious circle.

There is also another problem that is exemplified by this coordination game G in which the players’ interests partly diverge. “Pure” Nash equilibria are asymmetric—they favor one player (the one who dares) at the expense of

\textsuperscript{8}Guala uses other labels for the strategies in his example, which has to do with the particular story he tells to illustrate the game in question (see 44f): He describes competition for grazing areas between two tribes on the African savannah. But there is no need to retell this story here.
the other,\(^9\) while the symmetric “mixed” equilibrium offers each player a rather poor expected payoff, as it happens no higher than what she would be guaranteed if she directly went for \(C\).

Can both players do any better than that? Indeed, they can, in the presence of a correlation device. By observing the signals it sends, the players can coordinate their strategies. This requires that they interpret the device’s signals as instructions. There may be someone who operates the device (Guala calls such a person a “choreographer,” cf. 48). But the device might also be impersonal: There does not have to be anyone who stands behind its instructions to the players.

Suppose then that the correlation device specifies who is to dare and who is to chicken out. Its instruction is thus either \((C, D)\) or \((D, C)\), with probabilities for these instructions being \(p\) and \(1 - p\), respectively. Suppose the device is impartial: it does not favor any player. This means that \(p = \frac{1}{2}\). Thus the device might simply involve tossing a fair coin: if Heads, \((D, C)\) is prescribed; if Tails, \((C, D)\).

If both players choose to follow the device’s instruction, each player’s expected payoff is \(\frac{1}{2} \times 2 + \frac{1}{2} \times 1 = 1.5\). This is better than 1, which is the expected payoff to each player in the mixed Nash equilibrium. Furthermore, assuming that the other player is going to follow the instruction of the device, it is in each player’s interest to do likewise: She has no reason to deviate.

This is an example of what game theorists, following Robert Aumann (1974), call a correlated equilibrium.\(^{10}\) Real life abounds with interactions in which correlation devices play a crucial role. A standard example is an interaction between car drivers at a street crossing, with traffic lights functioning as a correlation device.

Now, Guala’s suggestion is that institutions essentially involve correlated equilibria. There is no institution yet if the players succeed in coordinating their strategies without help from a correlation device (cf. 51ff). In an institution, coordination is mediated by correlation. This implies that the player’s strategies can be described as conditional rules, or better, as conjunctions of conditional rules. In the Traffic Game, the strategy of the drivers is “If Green, drive; If Red, wait.” In the Chicken Game \(G\) we have considered earlier

---

\(^9\)A coordination game in which the players’ interests fully coincide might also have multiple pure Nash equilibria, but they are all symmetric. Indeed, even some games in which the players’ interests partly diverge might have symmetric Nash-solutions. Prisoners’ Dilemma is a case in point.

Contrast this correlated equilibrium with the mixed Nash equilibrium \((\frac{1}{2}(C, D), \frac{1}{2}(D, C))\)\(^{11}\). Row’s strategy is then “If Heads, D; If Tails, C,” while Column’s strategy is the opposite one: “If Heads, C; if Tails, D.”

In this particular case, the correlation device is impartial; it gives the same expected payoff to each player. But, needless to say, impartiality and fairness are not general features of institutions. The correlation devices on which institutions are based might favor some players at the other players’ expense. Thus, in Game G, \(p\) might be higher than \(\frac{1}{2}\), which would favor Row: High \(p\) makes it more probable that the device’s instruction will be \((D, C)\) rather than \((C, D)\). Or \(p\) might be lower than \(\frac{1}{2}\), which would favor Column.

Indeed, \(p\) might be 1, in which case the correlation device simply prescribes \((D, C)\). Or it might be 0, in which case the device prescribes \((C, D)\). Thus, every pure Nash equilibrium of the game is one of its possible correlated equilibria. The same applies, by the way, to every mixed Nash equilibrium as well: The correlation device might prescribe mixed strategies to the players. But there are, as we have just seen, correlated equilibria that are not reducible to Nash equilibria, either pure or mixed. In our example, those are the ones that assign to both \((D, C)\) and \((C, D)\) nonzero probabilities. There are infinitely many such probability distributions.

Let me now take up a side issue. In our example of a coin toss deciding who is to dare and who is to chicken out, each player knows what both players have been instructed to do. The instructions depend on a public event (the outcome of the coin toss), which makes them common knowledge. But it need not always be the case. While Guala does not mention this, some correlated equilibria might involve private signals: each player is instructed what to do, but she might not know what other players have been instructed to do. What is needed for correlation, though, is that each player knows the probabilities with which the correlation device issues different sets of instructions to the players.

Each such set of instructions is a profile of strategies, pure or mixed, with one strategy for each player. A probability distribution on such strategy profiles forms a correlated equilibrium if and only if it holds for each player \(i\) and

\(^{11}\text{Contrast this correlated equilibrium with the mixed Nash equilibrium }((\frac{1}{2}C, \frac{1}{2}D), (\frac{1}{2}C, \frac{1}{2}D)). \text{ While the former guarantees that the players will end up in either } (C, D) \text{ or } (D, C) \text{ and in particular will avoid the catastrophic } (D, D), \text{ the latter does not exclude this danger.}\)
Note that this general characterization of a correlated equilibrium implicitly assumes that instructions to each player are private. Intuitively, in a correlated equilibrium, a player’s expected payoff is maximized if she does what she is told to do. It is maximized given her prior knowledge of the probabilities with which the device issues different sets of instructions to the players. It is not required that doing what she is told would maximize her expected payoff if she learned what the other players have actually been instructed to do. While, in many cases, such additional information would not make any difference, in some cases it might, as illustrated by the example that follows.

If the instructions of the correlation device are common knowledge, the strategy profiles prescribed by the correlation device must always be Nash equilibria. Otherwise, some players would have reasons to deviate from the device’s instructions. But, if the instructions are private signals, this need no longer be the case. We can illustrate it in Table 2 with another version of Chicken, this time drawn from Aumann’s 1974 paper.

### Table 2. Aumann’s Chicken.

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>0, 0</td>
<td>7, 2</td>
</tr>
<tr>
<td>C</td>
<td>2, 7</td>
<td>6, 6</td>
</tr>
</tbody>
</table>

In this game, the 50-50 probability of $D, C$ and $C, D$ would give each player the expected payoff of $\frac{1}{2}(7) + \frac{1}{2}(2) = 4.5$. There is a better correlated equilibrium, though, one that offers higher expected payoffs to the players. But it requires that the instructions to each player are communicated to her in private. The idea is then that a correlation device picks out some strategy profile, with all the players knowing the probabilities with which each such profile can be picked out, and then it instructs each player in private what she is to do. Suppose both players know that the device can pick out $(C, C)$, $(D, C)$, or $(C, D)$, with probabilities $\frac{1}{2}$ for the first profile and $\frac{1}{4}$ for each of the other two. Note that one of these profiles, $(C, C)$, is not a Nash equilibrium. Still, it is easy to show that the

---

12Note that this general characterization of a correlated equilibrium implicitly assumes that instructions to each player are private. Intuitively, in a correlated equilibrium, a player’s expected payoff is maximized if she does what she is told to do. It is maximized given her prior knowledge of the probabilities with which the device issues different sets of instructions to the players. It is not required that doing what she is told would maximize her expected payoff if she learned what the other players have actually been instructed to do. While, in many cases, such additional information would not make any difference, in some cases it might, as illustrated by the example that follows.
probability distribution $\frac{1}{2}(C, C)$, $\frac{1}{4}(C, D)$, $\frac{1}{4}(D, C)$ forms a correlated equilibrium. Furthermore, as it is easy to calculate, it gives each player the expected payoff of $\frac{1}{2}(6) + \frac{1}{4}(2) + \frac{1}{4}(7) = 5.25$, which is considerably better than 4.5.

This equilibrium requires that players do not get the whole information: they are not to be told what the other players have been instructed to do. (If a player were told to do $C$ and informed that the other player has been told to do likewise, she would have reason to disobey the instruction and go for $D$ instead.) Nor should they always be able to deduce it from their knowledge of the probability distribution. In particular, if, say, Row is instructed to do $C$, she does not know what the other player has been told to do. Both $(C, C)$ and $(C, D)$ are assigned positive probabilities in this correlated equilibrium.

Now, I wonder whether Guala would agree that some institutions rely on correlation devices whose instructions are secret in this way. I would think so; full transparency does not seem to be essential to institutions. Perhaps not even to good institutions, if efficiency can be promoted by some secrecy.

This was just an aside. My main question is a different one: as we have seen, Guala introduces the idea of correlation in connection with the coordination problem that arises in games with multiple Nash equilibria. What I wonder is whether the introduction of correlated equilibria makes the coordination problem any easier. As far as I can tell, it does not. In a way, it makes this problem harder, since—as we have seen—the class of correlated equilibria of a given coordination game is much, much larger than the class of its Nash equilibria. Indeed, the former class is infinitely large and it includes the class as of Nash equilibria as its proper subset. Furthermore, as shown by Aumann’s Chicken, it is larger than the class of probability distributions on the set of Nash equilibria. It also admits probability distributions that allow for non-Nash strategy profiles.

Of course, Guala recognizes this problem of the multitude of correlated solutions. He suggests that the choice of a particular solution is a matter of
salience. The players coordinate on a correlated equilibrium that stands out in some way. And salience, at least to some extent, depends on the factors that lie outside abstract game theory. It is determined by the players’ psychology, by customs, historical precedents, and the like. It might also be brought about by the players themselves, by agreements or acculturation.

But there is no game-theoretic account of which correlations we naturally hook onto. Social ontology merges with social history, psychology, and biology at this point, and theoretical speculation must give way to empirically informed models. (50)

But then, one wonders, why has this move beyond Nash to correlated equilibria been made by Guala in the first place? Would the appeal to salience not be just as effective in explaining why the players manage to coordinate on a particular Nash equilibrium?

Indeed, as Guala recognizes (48), correlated equilibria of a given game, G, are themselves Nash equilibria of an augmented game G*—a game in which, for each player, the set of her strategies has been augmented by such conditional strategies as “If Heads, D; if Tails, C.” This strategy of a player, which is conditionalized on the workings of the correlation device, is in Nash equilibrium with the other player’s strategy “If Heads, D; if Tails, C.”

Guala also recognizes (49) that players might confront a choice between different correlation devices. There will normally be many such devices to choose between. And it is crucial for coordination that all the players regulate their actions by the same device.

Suppose that, in Guala’s Chicken (Game G), there are two correlation devices: a toss of a fair coin, issuing in either Heads or Tails, and a toss of a fair dice, issuing in either Odd or Even. Each player might then choose to regulate her action by one device or the other. Or she might instead directly opt for D or for C. We can represent the players’ quandary by the following augmented game G*(Table 3).

<table>
<thead>
<tr>
<th>Game G*</th>
<th>D</th>
<th>C</th>
<th>If Heads, C; if Tails, D</th>
<th>If Odd, C; if Even, D</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>0, 0</td>
<td>2, 1</td>
<td>1, ½</td>
<td>1, ½</td>
</tr>
<tr>
<td>C</td>
<td>1, 2</td>
<td>1, 1</td>
<td>1, 1.5</td>
<td>1, 1.5</td>
</tr>
<tr>
<td>If Heads, D; if Tails, C</td>
<td>½, 1</td>
<td>1.5, 1</td>
<td>1.5, 1.5</td>
<td>1, 1</td>
</tr>
<tr>
<td>If Odd, D; if Even, C</td>
<td>½, 1</td>
<td>1.5, 1</td>
<td>1, 1</td>
<td>1.5, 1.5</td>
</tr>
</tbody>
</table>
The Nash equilibria (in pure strategies) in G* are marked in bold. Some of the correlated equilibria of G are Nash equilibria in G*. Note that we can now consider various correlated equilibria of this new game G*, which will then in turn be Nash equilibria in a yet further augmented game G**, and so on. The players confront a coordination problem at each level. And the problem does not get any easier. If anything, it gets more and more difficult.

So, I wonder, what is the point of this move to correlated equilibria in the first place?

Guala answers this question as follows:

Clearly the theory of correlated equilibria does not solve the “mystery” of salience. It does, however, solve an important puzzle of social ontology. Recall the question we started from: are institutions rules, or equilibria of a game? We can now see that the answer is “both”: an institution may be considered as an equilibrium or as a rule of the game, depending on the perspective that one takes. (50)

How does the theory of correlated equilibria help us to see that the answer is “both”? Guala suggests that, from the point of view of an external observer, the institution is a correlated equilibrium: The players’ behavior “takes the form of a regularity that corresponds to the correlated equilibrium in G (or, which is equivalent, to the corresponding Nash equilibrium in G*).” On the other hand, from the player’s internal perspective, the institution is seen as a rule: “each strategy in this profile . . . takes the form of a rule that dictates each player what to do in the given circumstances.” The players “perceive the institution as a prescription” (50).

I must admit I am at a loss here. Why would the same two perspectives not be available without introducing correlated equilibria? Why can’t we say that from the point of view of an observer, the institution is a Nash equilibrium, while from the player’s perspective, her strategy takes the form of a prescriptive rule? Guala’s thought seems to be that institutional rules, in order to be perceived by the players as prescriptive, as something they are obligated to follow, have to be external in some sense. If I understand it correctly, the appearance of prescriptivity is possible according to him only if the rules are conditional on the workings of some external correlation device which the players themselves do not control. If this is what Guala means, then I wonder whether he is right. Why can unconditional rules not be perceived as prescriptive by the players? And, if they can, then why do institutions necessarily

---

14Cf. Guala, 51.
require correlated equilibria? Consider an institution such as a society’s morality. Some of its rules seem to be unconditional: “Don’t kill,” “Don’t torture.” And, even if we only focus on conditional moral rules, many of them do not appeal to correlation devices. Just think of a rule requiring us to make amends if we have hurt someone. It is conditional but does not refer to an external correlation device. So why do institutions have to be correlated equilibria? Why this move beyond Nash? It would be good to have it clarified.

When I tried to answer this question on my own, I thought of two possible responses.

One possibility might be that, in many coordination games, correlated equilibria offer better expected payoffs to the players. We have already seen this in Guala’s Chicken.

If both players are confident that the other player is going to follow this strategy [of conditionalizing her behavior on a toss of a fair coin], then they are better off tossing the coin. They would always coordinate on an efficient outcome . . . . (48)

So, the idea is that good institutions— institutions that promote the interests of their participants— can be expected to rely on correlation devices.

However, this response is not satisfactory. As we know, there are many institutions that are not good at all: they favor some of the participants at the expense of the others.

A possibly more promising suggestion, though I have not found any intimation that Guala had something like this in mind, is that coordination on a correlated equilibrium creates a causal linkage, an interdependence between players. This causal interdependence is mediated by the operation of the correlation device: its signals to the players are the common cause of the players’ actions. It might be plausibly argued, I suppose, that an institution cannot exist without causal interdependence between its members. If the players independently act in such a way that they end up in a Nash equilibrium, there is no institution yet.

However, even this response does not seem to be fully satisfactory, at least not as it stands. Even if some form of interdependence between the participants is required for the existence of an institution, does it have to be mediated by a correlation device? Can it not be more direct than that? I am not sure. Perhaps some correlation devices are necessary for causal interactions within institutions. But, again, it would be good to have it clarified.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Author Biography
Włodek Rabinowicz is senior professor of Practical Philosophy at Lund University and has numerous publications in value theory, moral philosophy, decision theory, and philosophical logic. He is an editor of *Theoria* and a former editor of *Economics and Philosophy* and *Philosophy and Phenomenological Research*. Rabinowicz is a long-term fellow of the Swedish Institute of Advanced Study and honorary professor at the Australian National University and University of York.