3-2013

Rule-Following as Coordination: A Game-Theoretic Approach

Giacomo Sillari
University of Pennsylvania

Recommended Citation

Sillari, G. (2013). Rule-Following as Coordination: A Game-Theoretic Approach. *Synthese, 190* (5), 871-890. http://dx.doi.org/10.1007/s11229-012-0190-z

Rule-Following as Coordination: A Game-Theoretic Approach

Abstract

Famously, Kripke has argued that the central portion of the Philosophical Investigations describes both a skeptical paradox and its skeptical solution. Solving the paradox involves the element of the *community*, which determines correctness conditions for rule-following behavior. What do such conditions precisely consist of? Is it accurate to say that there is no fact to the matter of rule following? How are the correctness conditions sustained in the community? My answers to these questions revolve around the idea (cf. P.I. §§198, 199) that a rule is followed insofar as a convention is in place. In particular, I consider the game-theoretic definition of convention offered by David Lewis and I show that it illuminates essential aspects of the communitarian understanding of rule-following.

Make the following experiment: *say* "It's cold here" and *mean* "It's warm here". Can you do it?

Ludwig Wittgenstein, *Philosophical Investigations*, 1953, §510.

I can't say "it's cold here" and mean "it's warm here"—at least, not without a little help from my friends.

David Lewis, *Convention*.

Keywords

coordination, rule-following, convention, Wittgenstein, David Lewis, common knowledge

Disciplines

Epistemology | Logic and Foundations of Mathematics | Metaphysics | Philosophy of Language | Philosophy of Science

Rule-following as coordination: A game-theoretic approach

Giacomo Sillari
University of Pennsylvania

> Make the following experiment: *say* "It's cold here" and *mean* "It's warm here". Can you do it?
> Ludwig Wittgenstein, *Philosophical Investigations*, §510.

> I can't say "it's cold here" and mean "it's warm here"—at least, not without a little help from my friends.
> David Lewis, *Convention*.

Abstract

Famously, Kripke has argued that the central portion of the Philosophical Investigations describes both a skeptical paradox and its skeptical solution. Solving the paradox involves the element of the *community*, which determines correctness conditions for rule-following behavior. What do such conditions precisely consist of? Is it accurate to say that there is no fact to the matter of rule following? How are the correctness conditions sustained in the community? My answers to these questions revolve around the idea (cf. P.I. §§ 198, 199) that a rule is followed insofar as a convention is in place. In particular, I consider the game-theoretic definition of convention offered by David Lewis and I show that it illuminates important elements of the communitarian understanding of rule-following[1].

0. Introduction

The slogan that "meaning is normative" (and the normativity of rule-following in general) is best understood in the context of *strategic* interaction in a community of

[1] Many thanks for useful commentaries and discussions to Peter Baumann, Cristina Bicchieri, Liz Camp, Ka-Yuan Cheng, Richard Eldridge, Francesco Guala, Simon Huttegger, Rohit Parikh, Jan Sprenger, Kevin Zollman, Brian Skyrms, and audiences at the 2008 Summer School Urrutia Elejalde on Social Norms, San

individuals. Famously, Kripke has argued in (Kripke 1982) that the central portion of the *Philosophical Investigations* describes both a skeptical paradox and its skeptical solution. Solving the paradox involves the element of the *community*, which determines conditions of assertability in the language, and conditions of correctness for rule-following behavior. A battery of argument is used to argue that meaning (or, in general, rule-following) cannot be explained by resorting to an individual's mental states, or her past use, or her dispositions. By exclusion, this indicates that no descriptive fact is constitutive of meaning and that "meaning is normative" or, for the purpose and scope of this article, that rule-following is a normative notion. The normativity of rule-following is related to the correctness conditions that hold in a community. Indeed, membership in the community depends on one's record of compliance[2] with the correctness conditions. But *what* do such conditions precisely consist of? *How* are they sustained in the community? And is it accurate to say that there is no *fact* to the matter of rule-following[3]?

The central thesis of this article is that the skeptical solution put forth by Kripke can be illuminated if looked at in the context of the *strategic* interaction taking place in a

Sebastian, the 31st International Wittgenstein Symposium in Kirchberg, Austria, the 83rd meeting of the Pacific APA in Vancouver, Canada, the Philosophy Department at Swarthmore College and the 37th meeting of the Society for Exact Philosophy in Alberta.

[2] Cf. (Kripke 1982: 91-2): "Any individual who claims to have mastered the concept of addition will be judged by the community to have done so if his particular responses agree with those of the community in enough cases […] An individual who passes such tests is admitted in the community as an adder; an individual who passes such test in enough other cases is admitted as a normal speaker of the language and member of the community."

[3] While Kripke himself uses the terms meaning and rule-following rather interchangeably in (Kripke 1982), in some cases the two terms need not be equivalent. In the following, I tackle the issue of rule-following and of the nature of correctness conditions, leaving the analysis of meaning to a companion paper.

community. Wittgenstein states that a rule can be followed only insofar as there is a habit, a social convention in place. Building on Wittgenstein's observation, we will read the skeptical paradox of the central portion of the Investigations, as well as Kripke's skeptical solution in light of David Lewis's definition of social convention in his seminal *Convention: A Philosophical* Study. Lewis offers his account in the context of the theory of games, as we do in this paper. The game-theoretic analysis allows us to better understand the role played by the community in the skeptical solution and, in particular, it will allow us to explain in greater detail the role played by some key notions in the Kripkean approach, as those of agreement, *Lebensform* and *blind action.*

The rest of this article is organized as follows. In section 1, I introduce Lewis's theory of convention and briefly discuss its normative content. In section 2, I address two objections to the opportunity of a Lewisian and game-theoretic approach to rule-following. In section 3, I interpret Wittgensteinian rule-following in terms of Lewis-conventions. In section 4, I discuss the relation of the notions of *Lebensform* and *common knowledge*, while section 5 will be concerned with Wittgensteinian blind action in the context of evolutionary game theory.

1. Convention as Coordination

Wittgenstein states (§§198, 199) that a rule is followed *insofar* as there exists a custom, a convention. Yet, as Bloor (1997: 27) points out, "[w]e need more than a generalized awareness of the importance of social processes: we need a specific understanding of what is meant by the word 'institution[4]'." I claim in this paper that Lewis's game-theoretical account of convention can answer such a need, and that in fact the idea of rule-following as participation in a social custom or institution is illuminated when looked at through the lens of David Lewis's theory of convention. Lewis argues in (Lewis 1969) that coordination games (situations of strategic interaction in which the interest of the players roughly coincide) underlie every instance of convention, in that a convention is a regularity in the solution (equilibrium) of recurrent coordination games. The agents participating in the convention conform to the regularity because they prefer conformity over non-conformity, conditional on other agents' conforming. They form the belief about other agents' conformity through some coordination device: explicitly—through agreement—or tacitly—because a certain action stands out as the one that most

[4] Shortly thereafter, Bloor (*ibidem*) complains that "positive ideas on the subject has been conspicuous by their absence," and proceeds to put forth his own positive notion of social institutions, drawing both from philosophy (Anscombe) and sociology (Barnes.) I'm highly sympathetic to Bloor's view, in which a social institution is defined as a "collective pattern of self-referring activity" (Bloor 1997: 33). While Bloor approvingly mentions Lewis's formal rendition of the Humean account of social convention, he does not emphasize the strategic element inherent in social conventions—the game-theoretic idea that an agent (and indeed the whole community) is *better off* coordinating her behavior with that of others on the prevailing social convention. But it is precisely this material element that, as I shall argue throughout this paper, gives explanatory strength to the game-theoretic understanding of social institutions and in so doing clarifies the notion of the normativity of rules.

likely (almost) everyone will pick. Such an action is *salient*[5] to the parties. In the case of a recurrent coordination problem, a special kind of salience—*precedent*—is at play.

For an instance of a coordination game involving salience as a coordination device, suppose that two friends are hiking a trail at a distance from one another. The first hiker reaches a bifurcation. She does not expect the hiker who is behind to be able to see which path she is going to take, hence she collects a few stones and sticks and improvises a signpost. The interest of the hikers coincides (they want to pick both the same path) and the situation is strategic, in that each hiker prefers one direction over another conditional on the choice of the other. The situation can be depicted by the following matrix in which the numbers in each cell correspond to the payoffs received by the players

		II	
		left	right
I	left	(1,1)	(0,0)
	right	(0,0)	(1,1)

The matrix represents the interaction described above, and inspection promptly reveals that it possesses two (pure[6]) equilibrium points (*R*,*R* and *L*,*L*), that is two combinations of

[5] The notion of *focal point* was introduced by Schelling (1960) as the explanation for successful coordination in informal experiments.

actions such that in each of them no player has an interest to change her action. Coordination on one equilibrium point rather than another (in the example, *R*,*R* rather than *L*,*L*) is achieved by using the signpost as a coordination device, the *salient* equilibrium being the one obtainable by following the direction indicated by the hand of the signpost. We could stretch the example further and imagine that the two friends find themselves *often* in the situation just described. Instead of the elaborate signpost procedure, over time they could rely on more expedite mechanisms to achieve coordination. They may even simply rely on past coordination and always take the path to the right, which has become salient as a coordinative outcome because of precedent. It has become a regularity.

Such a regularity, which in Lewis's account constitutes a convention[7], is sufficient for some degree of normativity to arise[8]. Indeed, in a community in which a certain custom is

---

[6] The game also possesses a (weak) equilibrium in *mixed* strategies. In particular, if player I plays left and right with probability ½, then any action of player II nets player II the same payoff as any other action. Similarly for player II playing left and right with probability ½. Hence neither player has (in a weak sense) an interest to deviate from such randomized strategies.

[7] Lewis's almost final (and sufficient for our purposes) definition of convention is the following (cf. Lewis 1969:58):

> A regularity *R* in the behavior of members of a population *P* when they are agents in a recurrent situation *S* is a *convention* if and only if it is true that, and it is common knowledge in *P* that, in any instance of *S* among members of *P*,
>
> (1) everyone conforms to *R*;
> (2) everyone expects everyone else to conform to *R*;
> (3) everyone prefers to conform to *R* on condition that the others do, since *R* is a coordination problem and uniform conformity to *R* is a coordination equilibrium in *S*.

in place—say, the custom of going by signposts—there is an equilibrium in the actions and beliefs of the agents involved such that the agents prefer conformity to the custom, provided that all other members in the community act according to the convention. If I do *not* go by sign-posts, or I go by them in a funny, abnormal way (for instance, going in the direction opposite to the one indicated, as mentioned, e.g., in §85) I act contrary both to my preferences—because I will not get where I intend to go—and to the preferences of other members of the community—because, say, I will end up being late, or not showing up at all. My reputation will suffer[9]. This indicates that, in general, parties to a convention feel the pressure to conform, to some degree. As Lewis puts it[10], "conventions

---

The notion of *common knowledge* was introduced in the philosophical literature by Lewis's essay. It indicates the state of affairs in which every agent in a group *G* of agents knows that *p*, everyone in *G* knows that everyone in *G* knows *p*, and so on. The notion of common knowledge has generated a vast literature in disparate fields, from logic to mathematics, to psychology and computer science. For a general overview, let me refer the interested reader to Vanderschraaf and Sillari (2007) and the references therein.

[8] Approaching the issue of normativity of conventions from the other endpoint, Gibbard (1994: 98-99) says: "How, then, do I explain accepting a norm? I explain it by placing it in a speculative psychology. Accepting a norm, I hypothesize, is a state of mind that is linked to a special kind of linguistically infused motivation or tendency. The tendency, roughly, is to do what the norm says. The psychic mechanisms that underlie this state have as a chief biological function coordination through discussion---with coordination taken in the broad, game-theoretic sense expounded by Thomas Schelling (1960: ch. 2)." Similarly, Gibbard (1990: 64) states that "[s]ystems of normative control in human beings, I am suggesting, are adapted to achieve interpersonal coordination. What might this mean? To answer this I sketch work of Thomas Schelling on rational coordination in pursuit of human goals, and John Maynard Smith's evolutionary analog of Schelling's theory." As Jason McKenzie Alexander (2008: 278) rightly points out, however, "[p]roblems of interpersonal coordination are certainly important for understanding human nature, but not all interpersonal decision problems are problems of coordination, even under the broad conception of coordination urged by Gibbard." I will expand on this issue in the next section.

[9] Thus, Lewis (1969: 99): "The poor opinion [other parties] will form of me, and their reproaches, punishment, and distrust are the unfavorable responses I have evoked by my failure to conform to the convention."

[10] But cf. Guala (2008) for a different interpretation of this quote.

are a kind of social norms." They have varying degrees of normative force, depending on how serious the consequences would be, were the convention to be broken[11].

2. Rules and Games

Before defending the claim that rule-following in the sense of Wittgenstein is best understood as a regularity in the solution of coordination problems—that is, as a convention in the sense of Lewis—I need to answer two methodological objections. First, in *Convention* Lewis does take into account rules (cf. Lewis 1969: 100-107) to conclude that it would be difficult to single out a sense of the word "rule" that agrees with his definition of convention. Second, it may seem that by using Lewis's game-theoretic approach in interpreting Wittgenstein's rule-following considerations, I am endorsing the view that Wittgensteinian language-games could or even should be understood exclusively in the context of game theory, while such context appears to be much too narrow to accommodate the cluster concept of "language-game." I respond to these objections below.

[11] Cristina Bicchieri (2006) distinguishes between conventions, descriptive and social norms. The distinction at work here is the one between descriptive and social norms. The dynamics of the (possible) acquisition of normative force by a descriptive norm is succinctly captured by Bicchieri (2006: 39): "what starts as a descriptive norm may in time become a stable social norm," and more extensively analyzed in Bicchieri (1993: ch. 6).

As for the first objection, Lewis points out in a section of *Convention* entitled "Rules" that "[w]e might be tempted to try distinguishing several senses on the word "rule," hoping that one of them would agree with my definition of convention" adding immediately thereafter that, however, "I doubt that the project would succeed[12]." This assessment follows the analysis of various circumstances indicating that in natural language there is no perfect overlap between so-called rules on the one hand and Lewis-conventions on the other. In particular, while *all* instances of convention can be thought of as (informal, tacit, unwritten, etc[13].) rules, not all "so-called" rules can be thought of as conventions. For instance, there are rules enforced so forcefully that one has an incentive to abide by them regardless of the behavior of others, hence the element of coordination is lacking and we cannot properly speak of a convention in the sense of Lewis. There are norms issued by some authority or power such that one has an incentive to obey them unless *everybody* else disobeys. Since the incentive to follow the rule is (almost) unconditional, these rule do not qualify as conventions in the sense of Lewis either. There are social or political obligations whose underlying strategic interaction is best represented by games of cooperation rather than by games of coordination, and hence again they are not conventions in the sense of Lewis[14]. There are many *so-called* rules

[12] Cf. Lewis (1969: 105.)
[13] Cf. Lewis (1969: 100, 105)
[14] A game of *cooperation* is a strategic interaction in which the socially optimal outcome (in which the sum total of the payoffs is maximized) requires the cooperation of both players, while individual rationality pulls each player toward non-cooperative strategies, leading to suboptimal outcomes. The paradigmatic example of a game of cooperation is the prisoner's dilemma. Each player can "cooperate" or "defect". From an individual player's point of view, the best outcome obtains when she defects while the other player cooperates; the second best when both players cooperate; the third best when both players defect and

that do not necessarily presuppose an underlying game-theoretic structure at all: this is the case of maxims, generalizations, laws of nature, hypothetical imperatives, and so on. On the other hand, "[i]t is harder to argue that some conventions are not naturally called rules[15]." And indeed it is *this* direction of the relation between conventions and rules that interests us here. In fact, a different way to state the claim of this article is to say that Wittgensteinian rule-following deals with situations identifiable *insofar as a there is a custom*. Thus, while not all rules are interpretable as Lewis-conventions, all rules pertinent to Wittgensteinian rule-following[16] involve a conventional element and hence can be analyzed as pertaining to situations in which individual preferences regarding their actions are conditional. Such situations are consistent with Lewis's analysis of convention in terms of coordination and in fact, as the rest of the article will show, *are* best understood as recurrent coordination problems.

To respond to the second objection, notice that the question crucial to the framework of this article is the following: Are we entitled to cast the rule-following considerations in a

---

the worst outcome obtains when she cooperates while the other player defects. The need for a more expansive use of game theory in modeling social phenomena and conventions was very early recognized by Ullmann-Margalit (1976). In addition to coordination, Sugden (1986) makes essential use of cooperation and of several other game-theoretic tools in modeling rational social interactions. Binmore's complex and bold theory of justice and the social contract (1994, 1998, 2005) has both games of coordination and cooperation as its basic building blocks. In the words of Jason McKenzie Alexander (2008): "[t]he key to our moral nature, rather, lies in the fact that we all face repeated interpersonal decision problems—of many types—in socially structured environments."

[15] Cf. Lewis (1969: 104.)

[16] Wittgenstein does not appear to be indiscriminate in his concerns with rules, cf. *Zettel* §320: "'cookery' is defined by its end, whereas 'speaking' is not. […] You cook badly if you are guided in your cooking by rules other than the right ones; but if you follow other rules than those of chess, you are *playing another game*."

game-theoretic account of convention? Now, the multifarious abundance of the remarks constituting the *Philosophical Investigations* is notoriously unsystematic. The vagueness of the family-resemblance notion of language-game is lost in the exactness of the mathematical definition of a game. For example, the process of inventing and changing language-games is often mentioned in the *Investigations*, yet in the game-theoretic approach we consider pre-existing games whose structure does not change upon repetition. Moreover, some kinds of linguistic interactions that Wittgenstein subsumes under the family-resemblance notion of language-games can difficultly, if at all, be analyzed in game-theoretic terms. However, there *is* a similarity that criss-crosses and overlaps throughout the examples[17] of §23 and elsewhere, and that is relevant to the game-theoretic interpretation I am proposing here. Most of the examples of language-games in §23 and elsewhere in the *Philosophical Investigations* function when immersed in a strategic interactive context[18]. Let me use a few examples to illustrate the idea. First, consider "guessing riddles." Guessing a riddle presupposes the existence of a framework in which an utterer uses a metaphoric and figurative language, or a pun, or semantical ambiguity to convey a definition through images that allow, roughly speaking, for a uniquely consistent semantical interpretation. Meanwhile, the audience attempts to

[17] Let me remind the reader of the variety of those examples: "Giving orders, and obeying them—Describing the appearance of an object, or giving its measurements […] Reporting an event—Speculating about an event—Forming and testing a hypothesis […] —Singing catches—Guessing riddles—Making a joke; telling it" and so on.

[18] If some (e.g. "singing catches") seem to resist game-theoretical analysis, for other items in the list of §23 a game-theoretic perspective sheds interesting light. Consider e.g. "[f]orming and testing a hypothesis." Bicchieri (1988) explicitly analyzes rules of scientific methodology as Lewisian conventions. More recently, Zamora-Bonilla (1999, 2006) argues that game theory can be fruitfully applied to understand essential aspects of the scientific enterprise.

produce such a unique consistent interpretation. One who is listening to a riddle (or reading it) and who does not understand it as the (ambiguous) definition of some thing or notion is equivocating and betraying the intentions of the author, and in so doing she makes the riddle, along with its meaning, vanish into thin air—or at least into uninteresting literal prose. Second, consider "reporting an event" and "speculating about an event." Again, the function of the audience is essential: Suppose that the governor has been arrested today. "The governor was arrested on corruption charges" and "The governor is guilty of corruption" are successful instances of reporting and speculating on the event only insofar as the audience recognizes the former utterance as declarative and factual, and the latter as hypothetical and possible. If an audience mistakes "The governor is guilty of corruption" for a factual, declarative utterance, the mismatch between the utterer's and the audience's understanding of the linguistic exchange is such that it ceases to qualify as a "speculating on an event" language-game. Lastly, consider the famous "builder-assistant" example of a language-game (§2). Here the builder asks the assistant for the stones necessary to the construction; the builder calls, for instance, "slab!" or "beam!" and the assistant brings him a slab or a beam. In this interaction, the builder observes a state of affairs (lack of a slab, say), emits a corresponding signal (the call "slab!"), and the assistant performs a corresponding action (brings a slab.) The strategic component should be apparent: the interaction between the builder and the assistant is successful if and only if slabs are brought when slabs are needed, beams are brought when beams are needed, etc. Thus, for the assistant bringing a slab is the correct action

given that the signal "slab!" has been sent to him because of a lack of slabs, and the signal "slab!" is the correct signal to send if bringing a slab is the assistant's correct response to it. This is in fact a problem of coordination of the same kind as those underlying Lewis's account of convention[19].

Even though the strategic element is essential for several instances of Wittgensteinian language-games, it remains true that large parts of the Wittgensteinian analysis are lost in a formal, game-theoretic account. For instance, the process of inventing and changing language-games is often mentioned in the *Investigations*, but in the game-theoretic approach we consider pre-existing games whose structure does not change upon repetition. As I have just pointed out, some kinds of linguistic interactions that Wittgenstein subsumes under the family-resemblance notion of language-games can difficultly, if at all, be analyzed in game-theoretic terms. It is no surprise that the 'cluster-concept' of language-game cannot be entirely covered by a game-theoretical account. As mentioned above, coverage is lost of some cases Wittgenstein draws our attention to. Other cases we can cover, losing some of details yet, to be sure, gaining in clarity. As I hope the rest of this paper will demonstrate, in a game-theoretic framework concepts that are key to the rule-following considerations can be more closely scrutinized and elements of the communitarian approach to rule-following better specified and illuminated.

[19] Cf. also §86, in which 'tables' are imagined that serve the purpose of linking given signals to appropriate action, with different systems of 'arrows' pointing from signals to actions emphasizing the arbitrariness of the interpretation of signals. In fact, the 'table' and the 'arrows' of §86 can be thought of as representing the 'receiver' portion of a Lewis-signaling system.

3. Rule-following as Coordination

The crucial sections concerning the "paradox" of rule-following revolve around §201. In §198 it is clarified that interpreting a rule is not sufficient to show what an agent is to do in order to follow the rule. All interpretations "hang in the air" (§201), possibly contradicting one another. They seem to be conceptually closer to *thinking* that one is obeying the rule rather than to actually *going* by the rule (§202.) The relation between the expression of a rule and the agent's action is perhaps established by a learning process (§198), indicating not only[20] that the connection between the expression of the rule and the action in accord to it is of a causal nature, but also that the existence of a convention is a necessary condition for the phenomenon of rule-following.

Thus, when judging whether an agent adheres to a rule, we cannot base our evaluation on interpretations (be them the agent's, or ours) of the rule. Rather, we need to observe the agent's action—or, better, her *inter*action with others. While an agent's *choice* of action may remain in some instances the endpoint of an interpreting process[21], action ceases to

[20] Following McDowell (1984: 360, n. 22,) I split the last paragraph of §198, assigning the first period to the interlocutor. Also, I take the "Nein; ich habe *auch*" of the last reply to the interlocutor as countering the "Aber damit hast du *nur*" (my emphases) in the interlocutor's line. Thus, the "nein" is not entirely adversative, but rather it can be read as "not only."

[21] However, the interpreting process cannot ultimately produce a justification for action. I shall say more about the issue of justification in the next section.

be an interpretation and manifests the only instance of rule-understanding that is liable to verification[22].

In his skeptical solution, Kripke maintains that, while many interpretations of a given rule may arise, there is (roughly speaking) only one correct way to abide by the rule. The correct application of the rule is determined by the community. In particular, the customary action is the action that accurately corresponds to the rule. The agent is supposed to do "what he is inclined to do" (Kripke 1982: 88) and his action is then assessed against the background of community practice. So, there is no logical link between rule and action, but rather a *psycho*logical link, validated (or countered) by the customary practice that have place in the community. One problem with this solution is that the paradox suggesting the impossibility of solipsistic rule-following applies also to the community. *How* is the customary action determined? *Why* is it so defined? As in the solipsistic case, communitarian interpretations of the rule based on past use are no sufficient grounds to answer such questions, since the rule is susceptible of a multiplicity of interpretations also when the task of determining rule-obeying behavior is left to the community rather than to its individual members. Communal dispositions do not provide firmer grounds[23]. Still, while "each of us […] calculates new addition problems, […] the

[22] Many agree on this matter, from both the individualistic and the communitarian sides: from the former, cf. for instance Baker and Hacker (1984;) from the latter, e.g., cf. Meredith William's argument about the *primacy of action* (Williams 1989: 183 ff.)

[23] This is a major objection against the view involving community's inclinations or dispositions. It recurs in many arguments by individualistic critics. Kripke himself sharpens his position by pointing out that the

community feels entitled to correct a deviant calculation" (Kripke 1982:111.) Where does such an entitlement come from? It is firmly grounded in the *practice* of the community, yet while Kripke (*ibidem*, 95-101) spells out the main ingredients of the skeptical solution (agreement, form of life and criteria,) it remains unclear why and how such elements succeed in bringing about a (practical) solution to the paradox. Again, it seems, no descriptive fact discriminates the practice of rule-following from failing to do so, and again we are lost in the "gulf between an order and its execution" (cf. §431).

Meredith Williams (1989, esp. ch. 6) forcefully defends the community view by putting forth an original reading of the rule-following considerations that goes beyond both Kripke's skeptical solution and various versions of individualism, most notably Baker and Hacker's "rules as internal relations." She stresses (Williams 1989: 169) that "[t]he normativity of rules is grounded in community agreement over time," although (*ibidem*: 177) "the community is not required in order to police the actions and judgments of all members, but in order to sustain the articulated structure within which understanding and

---

theory that one follows a rule insofar as she acts in the way most people in the community do "would be a social, or community-wide, version of the dispositional theory, and would be open to at least some of the same criticism as the original form." (Kripke 1982: 111). Boghossian (1989: 173) criticizes communal dispositionalism as a solution of the skeptical paradox. Blackburn (1984: §3) focuses on a similar point to question the skeptical solution, cf., e.g., "[t]he community is as much at a loss to identify the fugitive fact as the individual was." And, further on in the same section: "the skeptic who won against the private individual looks equally set to win against a community which has the benefit of mutual support." One of McGinn's (1984) arguments against the communitarian view is that, against what Wittgenstein says in the *Investigations*, it endorses the idea of rule-following as an interpretive process in which the interpretation is yielded by society rather than by the individual. If we cannot resort to a community's inclinations or disposition, then we need to look at the community's *practice*. I maintain that that is best done in the context of strategic interaction.

judging can occur and against which error and mistake can be discerned." The articulated structure of society forces us to ask questions different from the ones made moot by the skeptical paradox. Not 'what is the correct course of action with respect to a given rule?,' but rather 'what is the connection between a given rule and action?' Not 'what is the grammatically correct way to proceed?,' but rather 'what grammar is immanent in our practices?' The community does not provide its members with *standards* against which they can evaluate whether actions are in accord or contrary to rules. However, the community is structured in a way that sustains constancy of practice over time, that is *agreement* in the way we (generally) follow rules. In this sense, society in its articulated structure does not police the actions of individuals—that is, does not say outright what course of action is correct with respect to a rule, checking the rule against a given standard—but merely sustains an immanent "grammar" of our societal practices—that is, creates conditions such that coordinative action can be performed. Can we say more about the articulated structure of society that is necessary to discern correct and incorrect applications of a rule?

To answer this question, let me elaborate my reading of Wittgenstein following the sign-post example. As I understand it, the rule-following phenomenon presented in the *Philosophical Investigations* can be analyzed in three elements: the *expression* of a rule,

the *interpretation* of a rule, and the actual rule-*following*[24]. Consider again the example of a signpost. *Firstly*, the *expression* of the rule is the physical instance of the sign-post. The sign-post is merely a piece of wood that "*by itself* seems dead." (§432). The existence of a convention implies that a recurrent coordination game be recognizable. Thus, 'going by a signpost' happens only insofar as a strategic interaction involving the use of signposts is recognizable. The expression of a rule (psychologically) determines the strategic interaction related to the rule. The expression of a rule makes it explicit that a strategic interaction is going to take place, and so, in a sense, we could say that it determines that there will be interacting *agents*. *Secondly*, agents *interpret* the rule. In the case of going by sign-posts, let us consider agents who have, for simplicity, two interpretations: going in the direction indicated by the arrow (say, right), or going in the opposite direction[25]. Recognizing the structure of the strategic interaction related to the expression of a rule pertains to the element of rule-interpretation, that is providing strategies on how to follow the rule. The skeptical paradox stems from the interpretations "hanging in the air along with what they interpret." Thus, when interpreting a rule we come up with a series of possible actions (yielded by several of the vast number of possible interpretations). In the game-theoretic understanding of rule-following, each of the many interpretations of the expression is a possible *action* of the game induced by the rule. Even if we were to carry

[24] For a similar exegesis, cf. Arrington (2001).
[25] Of course there is an infinite number of possible strategic interactions that can be associated with any rule-expression. This implies that the resulting coordination game is arbitrarily large.

out the interpretive process[26] and decide that a certain action were the correct application of the rule, §202 reminds us that no action gains special support by any given interpretation, since thinking one is following a rule is not the same as following a rule. Thus, *thirdly and lastly*, the actual rule-*following* takes place only in the instant the agent actually moves away from the sign-post and goes right, or left. This phase represents the agent's non-interpretive grasp of the rule (§201) and constitutes the only portion of the process described here that can be evaluated. If the agents' choices strategically match the choice of other(s) and the profile of actions constitutes the regular solution (equilibrium) of the coordination game, then the agents are in fact following a rule and receiving a positive *payoff*. The following table is a synthesis of the three aspects of rule-following, along with their game-theoretic counterparts:

[26] As we shall see more closely in the next section, precedent and strategic reasoning, pattern projection and common knowledge thereof may all be thought of as parts of a deliberation or of a justification that selects one action over another (say, going right rather than left.) It remains true that in actuality, deliberation and justification often leave place to automatic behaviors that are blind (§219) and without justification (§269.) This does not mean that deliberation and justification should not be part of the analysis, for, among others, the three reasons listed below. Firstly, there are cases of conventions that are yet not well established (that is, rules that preserve some ambiguity as to what the correct response might be), and hence leave room for *thinking*: coming up with an interpretation and deliberate whether our interpretation is correct. Since the context is interactive and strategic, an essential part of my deliberation will be my belief about your action, and my belief about your belief about my action, and so on (cf. *passim*, and Sillari 2005 for a discussion of higher-order iterated beliefs in the context of conventions). Secondly, there are cases in which a well-established convention *fails* and the automatic action strikes trouble. In some of such cases, especially if recurrent, the agent will be interested in checking whether the action automatically performed matches the preferences and expectation that she may ascribe to other agents in the community. Thirdly, there are cases in which we are to *evaluate* the performance of others. To use the example of Brandom (1983: 643-645), the "parrot trained reliably to say 'It's getting warmer' only as the temperature climbs past 80 degrees never succeeds in asserting that it is getting warmer". Its behavior coordinates with ours, but does not match the preferences and expectations that can justify it. In conclusion, while conventional action and rule-following behavior is by and large an issue of automatic behavior (cf. Bicchieri 2006: ch. 2), a *rational reconstruction* of the system of preferences and beliefs involved in choice and motive is not unwarranted.

Expression	Interpretation	Action
Recognizing a strategic interaction	determining viable actions	material outcome
(players)	(strategies)	(payoffs)

I believe that the game-theoretic structure just described capture the essence of William's 'articulated structure'. In the game-theoretic context, community agreement is a special kind of agreement in preferences and beliefs that fulfills the definition of social convention. Agreement in preferences and beliefs (Williams's "harmony of society") is not the same as a standard of correctness with respect to given rules, but represents the articulated—that is: *strategic*—setting in which and through which we can evaluate action. The strategic setting does not offer any way to extrapolate standards of correctness. It describes the social interaction that makes rule-following agents better off, hence sustaining the constancy of societal practice. It does not provide agents with written-in-stone standards about what they are to do, but makes sure that *if* they think others will perform action *a* in response to rule *r*, then they will be better off performing action *a* as well. Indeed, if we interpret rule-following as *regular coordination equilibrium* in recurrent coordination problems[27], then there *is* a clear and compelling fact to the matter of what "going by the rule" consists of. In particular, individuals who go by

[27] Notice that I am not requiring, here, that the equilibrium be identifiable *ex ante* by members of the community. What needs to be recognized by the agents is that there is a strategic interaction in place. That is, the satisfaction of my preferences does not depend straightforwardly on my action, but also, and essentially, on what you are going to do.

the rule net a higher payoff than individuals who fail to obey the rule[28]. The mutually beneficial outcome is after all the *purpose* of the sign-post (thought of, *a la Lewis*, as a coordination device,) and in this sense we can say, with Wittgenstein (§87), that "[t]he sign-post is in order—if, under normal circumstances, it fulfills its purpose." That is, if we manage to coordinate, perhaps even efficiently, our behavior. Moreover, even if the community, as Williams explains, need not police the activities of its members at all times, transgressing the rule may come at a price, both for the transgressor and for the agents who are interacting with him. Non-conformative behavior thus may in these cases end up being sanctioned (possibly and eventually with expulsion from the community), while conformative behavior perpetuates itself, since it is based on the agreement to act according to given rules. In this sense, agreement is the agreement in preferences and beliefs that support a specific equilibrium in the recurrent coordination game, while the normativity of rules may originates from it.

Thus, the "little help" needed by Lewis from his friends in the answer to the challenge of §510 reported in the epigraph consists then in their *agreeing* to different preferences and

[28] Indeed it sometimes seems that critics of the communitarian solution neglect to consider the pragmatic consequences of rule-following, that is, in the game-theoretic setting I privilege here, the payoffs for coordinating behavior as opposed to the payoffs for failing to do so. As Rohit Parikh (2002) emphasizes, Wittgenstein himself does not always delve on the subject: for instance, in the famous "five apples" example of §1, the passing hands of money from the shopper to the shopkeeper is elegantly ignored.

beliefs, switching in so doing from one solution of a recurrent coordination game to another[29]. Consider §224:

> The word "agreement" and the word "rule" are *related* to one another, they are cousins. If I teach anyone the use of the one word, he learns the use of the other with it.

I believe that the view expressed in this section captures the sense in which "agreement" and "rule" are related: A custom—and hence a rule—does not hold without an agreement in preferences and beliefs—and hence in coordinative, conventional actions—on part of the members of the community. In other words, if an *agreement* is in place, parties to it now prefer to conform to the agreed upon behavior, and believe that all share such a preference (and belief). But mutual preferences and beliefs of this kind form a convention, which has some degree of normative force, hence is a *rule*. On the other hand, if a *rule* is in place, parties to it relinquish their unconditional preferences and conditionally prefer to conform to the rule. They do in fact have reason to conform insofar as they believe that others share such preferences and beliefs, i.e. insofar as an *agreement*, as described above, is in place.

[29] Hacker (1996: 201) points out that in an *ironic* context one does say "it's cold" to mean "it's very hot." In fact, we can (crudely) game-theoretically account for irony using the following signaling system: player *i* feels cold, she ironically sends the signal "it's warm in here" expecting the audience to catch the irony and achieve coordination by shutting the window, rather than take the expression literally and fail to coordinate by bringing a sweater. The 'ironic equilibrium' of this simple 2x2x2 signaling game lies with the *anti*-signaling system (cf. Skyrms 1996: ch. 5) of the signaling game.

So, a rule is a mutual (tacit) agreement to act in accordance to a regularity of past behavior. The "paradox," as the second half of §201 explains, turns out to be a misunderstanding: the myriad possible interpretations of the rule are eliminated by the existing institution, and we can say that an act is "obeying the rule" only with regard to the one action consistent with the existing regularity. But our approach, so far, has been descriptive: I argued that the social structure sustaining rule-following in the sense of Wittgenstein is a strategic structure sustaining coordination over time. What happens if the regularity in coordination is broken? The difficult question of *justifying* a coordination equilibrium arises, cf. §206: "Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way. But what if one person reacts in one way and another in another to the order and the training? Which one is right?"

4. Induction and Justification

Consider again the example of coordination from section 1. The two friends have invariably managed to successfully coordinate on many instances, and once again during their hike they find themselves in the same coordination problem. Except that this time the leading hiker leaves a signpost indicating right and goes *left*. Their failure to coordinate reveals that, in a two-by-two interaction as the one we are considering here, one of the two, and only one of them, is not following the rule. But *who* is right?

In the context of Lewis's theory of convention, the individual is right whose behavior conforms to *precedent*. Without reliance on precedent, no conventional strategic interaction in the sense of Lewis is possible and, as I have argued in the previous section, without the context of strategic interaction the community is in no better position than the individual in providing a determination as to which course of action is in accord with the rule. However, a skeptical paradox[30] resurfaces in the game-theoretic account of convention. The paradox parallels the one identified by Kripke's in the *Investigations*. It is now yielded by the multiplicity of patterns that can be inductively projected from past conformity onto the coordination problem at hand. Thus, Lewis's answer that *precedent* works as a coordination device among the players has no teeth when considered by someone who remains sensitive to skeptical arguments, since the notion of precedent is susceptible to a multiplicity of interpretations just the same as the notion of rule is. In the more recent literature on this issue (cf. Skyrms 1996, Cubitt and Sugden 2003, Sillari 2005, Rescoria 2007) two main stances on the issue emerge. Brian Skyrms, on the one hand, emphasizes that in the case of symmetric coordination games the emergence of a steady convention is a "moral certainty", although *which* convention will emerge "is a matter of chance." (Skyrms 1996: 93). We shall see the relevance of Skyrms's argument in the next section. Cubitt and Sugden, on the other hand, focus on the role of *salience* and *inductive standards*. For them, a given state of affair (history of past play) provides

[30] Cf. for instance the analysis in Cubitt and Sugden (2003: sections 6 and 7). The link between pattern-projection in Lewis-convention and Wittgensteinian rule-following is highlighted also in Bardsley and Sugden (2006).

evidence from which players can extrapolate beliefs about future action. In particular, parties to a convention project a given pattern (call it *precedent*) because that pattern is *salient* to them. They are confident that (nearly) everyone is making the same inductive inference because inductive standards are shared in the population. The inductive reasoning based on precedent yields common knowledge (or, more precisely, common *reason to believe*) that parties to the convention will conform to it in the next instance of the coordination problem. Salience—that is, the inductive projectibility of a given pattern as precedent—generates the regularity that constitutes a Lewis-convention. In turn, conventions change over time, and our notion of salience changes with them. What counts as precedent today may not count as precedent tomorrow, when the current convention will have morphed into a different one.

There is no deductive, infallible passage from past to future conformity, no "hardness of the logical must." Rather, between the expression of a rule and the agent's action there is a causal connection. Such a connection may be based on training, as the interlocutor suggests in §198 (my emphasis):

> What sort of connexion is there here?—Well *perhaps* this one: I have been trained to react to this sign in a particular way, and now I do so react to it. [—]But that is only to give a causal connexion […]

Or it can be revealed through inductive reasoning, invoked within rule-following by the notion of convention and hence of precedent. Lewis was aware of the importance of inductive reasoning in his theory of convention[31], and in describing the process that gives rise to the infinite series of replication of one's partner's expectations he notices that they need "mutual ascription of some common inductive standards and background information, rationality, mutual ascription of rationality, and so on[32]." Such inductive standards need to be common (and to be common knowledge as well) for Lewis's argument to go through. The commonality of our everyday inductive practice relies of the commonality of those standards, but there is no explanation available for the commonality of inductive standards. Similarly, Wittgenstein warns us that "the standard [of good grounds] has no grounds!" (§482.) It simply is the precondition of there being a certain language-game that we all share common standards, that we all use the same *idiom of life* rooted in our common *Lebensform*.

These considerations suggest to understand the rich and profound notion of *Lebensform* as containing those grounding inductive standards that are a fundamental part of Lewisian conventions. Agreement in form of life then would entail agreement in inductive standards which in turn brings about agreement in beliefs, expectations and preferences about one another's conformity to precedent. The system of concordant beliefs about

[31] For a detailed formal reconstruction and discussion of the inductive processes implicit in Lewis's account of convention, cf. Cubitt and Sugden (2003).
[32] Cf. Lewis (1969: 56-7.)

each other conformity—the strategic structure of society—*inductively stem* from the fundamental agreement in form of life. In turn, from our concordant beliefs and conditional preferences stem our conventions and customs, and our common knowledge thereof, hence our capacity to reach a consensus in *actions*, that is, to follow a rule.

5. Precedent and Blind Action

Perhaps common knowledge is not necessary at all in order to maintain a social convention, and the form of life can be shared by agents unaware of sharing it. A problem with this account is that inductive reasoning generates *common knowledge* which is, by itself, insufficient to provide a justification of conformative behavior. Indeed, as Margaret Gilbert tersely points out in (Gilbert 1990), in Lewis's account of convention practical rationality does not yield any *justification* to act in conformity to precedent. Common knowledge of rationality and of past conformity is insufficient to deductively establish conformity in the case at hand[33]. In a sense, the notion of precedent lies at bedrock, where the spade is turned (§ 217) and one acts blindly (§ 219) conforming to the convention and obeying the rule by habit, or because of a psychological tendency[34] to be a 'conformist'. The connection between rule and action can become so entrenched to make rule-following behavior automatic, *blind*. Wittgenstein speaks of "blind action," Gilbert

[33] On the other hand, Weirich (2007) has recently argued that two rational agents playing a coordination game can bring about reasons to act coordinatively by forming foreseeable intentions to perform such acts.
[34] On the tendency of agents` to be 'conformist' in recurrent coordination games, cf. the fascinating conclusions drawn from experimental studies in Guala (2008).

speaks of an "a-rational tendency." For McDowell (1984), understanding is precarious and contingent, for there is no guarantee that my current grasping of a concept will continue to function tomorrow as well[35]. Williams (1989:169) points out that once an individual has come to master a given rule, she has acquired a *second nature*, having reached bedrock and acting blindly without ultimate justification.

In evolutionary game theory, rationality is an emergent property of the behavior of a-rational individuals programmed to follow given strategies. In this sense, such models are well suited to account for Wittgensteinian blind action. In Skyrms's evolutionary approach, the emergence and sustenance of stable social conventions is guaranteed by the dynamics of the model. In this paragraph we offer a version of the argument to model Kripke's example of addition, developing the suggestion (cf. Skyrms 1996, ch. 5) that evolutionary models of convention can used to defuse the skeptical paradox.

In evolutionary game theory, we consider a *population* of individuals. It is assumed that pairs of individuals meet at random and interact in a game. It is assumed also that each individual is "programmed" to choose a given strategy. Outcomes are assigned a payoff that represents not the utility, but rather the *fitness* received by each player. Individuals who net higher payoffs are more adaptive, and their proportion within the total population

[35] Cf. McDowell ([1984] 2002: 70): "at any time in the future my interlocutor's use of the expression in question may simply stop conforming to the pattern that I expect." And, *infra*, "[m]y right to claim to understand him is precarious, in that nothing but a tissue of contingencies stands in the way of my losing it."

will increase. The modeler then chooses a dynamics for the evolution of the population, and observes the behavior of the system. The *replicator dynamics* is often used, in which the growth of the subpopulation choosing, say, strategy *S* is relative to the difference between the fitness of *S* and the average fitness for the entire population. If the difference is positive, *S* entails an evolutionary advantage and number of individuals programmed to play *S* will proportionally increase. If the difference is negative, *S* is evolutionarily disadvantageous and the proportion of individuals playing *S* will decrease. The replicator dynamics is often used to model both genetic and *cultural* evolution, which is the one of interest here.

What happens when a population of individuals finds itself interacting in the game of figure 1? The dynamics has three equilibrium points: one in which the entire population plays *R*, another in which the entire population plays *L*, and a third in which half of the population plays *R* and the other half plays *L*. The first two equilibria, however, are *stable*: if the system is perturbed by any small deviation (for instance, a random mutation, a mistake in performing a given action, etc.), the dynamics will promptly bring it back to the equilibrium. Consider the case in which the entire population plays *L* and suppose that a group of mutants plays *R* instead. Individuals playing *L* will almost invariably meet other individuals playing *L* (since by assumption almost the entire population is playing *L*, and matching is random), while seldom meeting individuals playing *R*. Similarly for

individuals playing *R*. But the payoff for (*L*,*L*) is 1, while the payoff for (*R*,*L*) is 0, therefore the difference between the fitness of those who play *L* and the average fitness of the population will be positive (albeit small if the perturbation of the equilibrium is also small), while the difference between the fitness of those individuals who play *D* and the average fitness for the population will be close to -1. Hence the replicator dynamics will reinstate the "all *L*" equilibrium when perturbed. Similarly, of course, for the "all *R*" equilibrium. The case of the equilibrium in which the population is evenly divided between *R*-players and *L*-players is different. Although it is an equilibrium (the difference between the fitness of individual strategies and the average fitness of the population is 0), it is not a stable equilibrium. If the proportion of individuals playing, say, *L* increases no matter how minimally, *L* becomes more fit than *R* and the number of *L*-players will continue to grow until the entire population will be playing *L*. Similarly, the entire population will end up playing *R* if the unstable equilibrium is perturbed (it does not matter how much) towards *R*. "All *L*" has thus a *basin of attraction* (the states in which the dynamics leads towards "all *L*") containing all states in which more than 50 of the population plays *L* and similarly for "all *D*". We can illustrate this in the following diagram:

All *L* ←←←←←●→→→→→All *R*

Figure 2

where the arrows indicate the direction of genetic drift and the unstable equilibrium point in which the population is divided equally between *L*-players and *R*-players is indicated by the symbol ●.

Differential reproduction leads invariably to the fixation of a stable equilibrium, provided that we allow, as it is reasonable, for random small perturbations. In this kind of model, agents have no beliefs (nor of course common beliefs), they do not carry out any deductive or inductive reasoning, and hence it is not necessary that they share any inductive standard. Once the dynamics has established one of the equilibria as a *convention*, agents conform their behavior to precedent and the do so *blindly*. It is of course unreasonable to maintain that the idealized assumptions of the replicator dynamics are apt descriptions of human social behavior, yet these models show that using precedent as a coordination device can be justified by an evolutionary argument overriding the skeptical paradox—the salience of precedent is, as it were, *naturalized*, given that we are willing to accept the replicator dynamics with its assumptions as acceptable approximations of our cultural evolution[36].

Models based on the replicator dynamics are cogent when used to explain the emergence of a given behavior, once the material element inherent in social action is taken into

[36] Against this idea, cf. especially Sugden (2001).

account. Consider for instance the traditional example of *quaddition*[37] from Kripke (1982). An agent who behaves as a quadder in a population of adders is not going to fare very well. On the other hand, a quadder in a population of quadders is not going to do as bad. A population of quadders, however, is arguably not going to do as well as a population of adders. The inability to perform additions more complex than 56 + 56 would seriously hinder the potential for scientific and social progress of the population. A group of a few 'mutant' adders will end up taking over a population made mostly of quadders[38].

Thus, if we represent 'adding' and 'quadding' behavior with *A* and *Q* respectively we have the following symmetric ranking of strategy profiles: (*A*,*A*) > (*Q*,*Q*) > (*A*,*Q*) = (*Q*,*A*), giving rise to the so-called hi-lo coordination game:

	Add	*Quadd*
Add	(99,99)	(0,0)
Quadd	(0,0)	(1,1)

[37] It is the famous rendition by Kripke of the skeptical paradox supposedly recoverable in the *Investigations* (cf. Kripke 1982:8-9): *quaddition* is an operation between two numbers *x* and *y* such that it yields as a result *x+y* if the two numbers are less than 57, and 5 otherwise. If an individual has never summed up numbers greater than 57, then the rules of addition and quaddition are practically indistinguishable, hence the skeptical paradox.

[38] Someone could use considerations as this one to support an individualistic view of rules as based on an 'internal grammar' (a la Backer and Hacker 1984). Of course—would the objector go—adders would take over a quadding society: after all, when it comes to *adding numbers*, adders are right and quadders are wrong. However, the reason why adders take over in the context of my argument has less to do with the fact that quadders are 'wrong' than with the fact that quadders are *worse off*. Recall that, in fact, my considerations are made in the context of a skeptical solution.

Figure 3.

The game is a coordination game of the same kind of the one introduced in section 1, except that here the payoff for coordination of the two pure strategy (*A*,*A*) and (*Q*,*Q*) equilibria differ. As in the game of section 1, the hi-lo game also has an equilibrium in mixed strategies. The probability of playing *A* in the mixed equilibrium depends on the size of the payoffs and in the game depicted in figure 3 is (.99 *A*; .01 *Q*). We can interpret the game as played among individuals in a population *P*, and analyze it evolutionarily. In the evolutionary analysis, we have two *stable* states corresponding to two (strict) pure strategy equilibria (*A*,*A*) and (*Q*,*Q*) of the hi-lo game. In such states, all individuals in *P* play *A* and *Q*, respectively. There also is one unstable state in which both *A* and *Q* are played, corresponding to the mixed strategy equilibrium of the hi-lo game. In this state 1% of the population play *Q* while 99% of the population play *A*. Each stable state has a basin of attraction immediately to the left and right of the unstable equilibrium, as represented in the following diagram:

All *Q* ←●→→→→→→→→→→All *A*

Figure 4

The basins meet at the mixed equilibrium point. Hence, the size of the basin of attraction of the inferior equilibrium (*Q*,*Q*) depends on the magnitude of the distance between the

fitness yielded by (*A*,*A*) and the fitness yielded by (*Q*,*Q*). If we assume that a population of quadders finds itself at a very strong disadvantage vis-à-vis a population of adders (that is, the payoff for (*Q*,*Q*) is much lower than the payoff for (*A*,*A*)), then the likelihood that quadders will take over the entire population remains low. Even though it is possible to imagine a community in which everyone is a quadder, it will however be precarious: a sufficiently strong perturbation can rather easily push the population in the basin of attraction of the more efficient adding equilibrium. If, as Skyrms points out[39], in the case of symmetric coordination games as the one in figure 1 the emergence of a convention is a "moral certainty" yet which convention is selected remains largely a "matter of chance," in the case of the hi-lo game, which captures essential features of the Kripkean example, chance plays a lesser role, since the basin of attraction of the efficient equilibrium is so much greater.

7. Conclusion

Game theory sheds new light on the notoriously obscure pages of the *Investigations* dealing with rule-following. Taking at face value Wittgenstein's indication that following a rule requires that a convention be in place, I have used David Lewis's game-theoretic account of convention to clarify how rule-following presupposes agreement an coordination in a community. In so doing, the role played by the community is made more perspicuous, and in particular we have seen that the *strategic* component is crucial

[39] Cf. Skyrms 1996:93.

of a full understanding of rule-following. Game theory and the Lewisian analysis of social conventions shed light also on two notions related to rule-following. The notion of *Lebensform* is illuminated if looked at next to the technical notion of *common knowledge*, and the notion *blind action* is clarified in the evolutionary approach. As I have already stated above, I am not claiming that game theory can cover all subtle nuances in Wittgenstein's notion of language-game, and neither I claim that hard interpretative issues (for instance that of solipsistic vs. communitarian reading of rule-following) can be settled by game theory once and for all. However, I do believe that I have singled out a group of notions in the *Investigations* which find precise counterparts in normal game-theoretic ones. Finally, if my analysis does not of course purport to be historical in character, still it highlights that the later Wittgenstein already contains seeds of a philosophy of social sciences that has found voice first in David Lewis's seminal study and that, today, continues to grow at the intersection of philosophy and game theory.

Bibliography

Arrington, Robert L. (2001) "Following a Rule," in *Wittgenstein: A Critical Reader*, Hans-Johann Glock, ed., Oxford: Blackwell

Baker, Gary, and Hacker, P. M. S. (1986) *Skepticism, Rules and language*, Oxford: Blackwell

Bardsley, Nicholas and Robert Sugden (2006) "Human Nature and Sociality in Economics," in *Handbook of The Economics of Giving, Reciprocity and Altruism*, eds. Serge-Christophe Kolm and Jean Mercier Ythier, North Holland, Amsterdam, pp. 731-768.

Bicchieri, Cristina (1988) "Methodological Rules as Conventions," *Philosophy of the Social Sciences* 18:477-95

Bicchieri, Cristina (2006) *Grammar of Society*, Cambridge: Cambridge University Press

Bicchieri, Cristina (1993) *Rationality and Coordination*, Cambridge: Cambridge University Press

Blackburn, Simon (1984) "The Individual Strikes Back," *Synthese* 58(3):281-301

Brandom, Robert (1983) "Asserting," *Noûs* 17 (4):637-650

Bloor, David (1997) *Wittgenstein, Rules and Institutions*, London and New York: Routledge

Boghossian (1989) "The Rule-Following Considerations," *Mind*, 98 (1989), pp. 507-49

Cubitt, Robin and Robert Sugden (2003) "Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory," *Economics and Philosophy* 19:175-210

Gibbard, Allan (1994) "Meaning and Normativity," *Philosophical Issues*, pp. 95-115

Gibbard, Allan (1990) *Wise Choices, Apt Feelings*, Cambridge, Mass.: Harvard University Press

Gilbert, Margaret (1989) "Rationality and Salience,", *Philosophical Studies*, 57:61-77

Guala, Francesco (2008) "Are There Lewis Conventions?," working paper, University of Exeter

Hacker, P. M. S. (2000) *Wittgenstein: Mind and Will, Analytical Commentary on the Philosophical Investigations, vol. 4*, Oxford: Blackwell

Hacking, Ian (1993) "On Kripke's' and Goodman's Uses of 'Grue,'" *Philosophy* 68(265), 269-295

Kripke, Saul (1982) *Wittgenstein on Rules and Private Language*, Cambridge, Mass.: Harvard University Press

Lewis, David (1969) *Convention*: *A Philosophical Study*, Cambridge, Mass.: Harvard University Press

McKenzie Alexander, Jason (2008) *The Structural Evolution of Morality*, Cambridge, Cambridge University Press

McDowell, John (1984) "Wittgenstein on Following a Rule," *Synthese*, 58(3):325-64,

McGinn, Colin (1984) *Wittgenstein on Meaning: An Interpretation and Evaluation*, Oxford: Blackwell
Miller, Alexander and Crispin Wright, eds., (2002) *Rule-Following and Meaning*, McGill-Queen's, Montreal
Parikh, Rohit (2002) "Social Software," *Synthese*, 132(3):187-211
Rescoria (2008) "Convention," in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/convention
Schelling, Thomas (1960) *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press
Sillari, Giacomo (2005) "A Logical Framework for Convention," *Synthese*, 147(2):379-400
Sillari, Giacomo (2008) "Common Knowledge and Convention," *Topoi*, 27(1-2):29-39
Skyrms, Brian (1996) *Evolution of the Social Contract*, Cambridge: Cambridge University Press
Vanderschraaf, Peter and Giacomo Sillari (2007) "Common Knowledge," in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/common-knowledge
Weirich, Paul (2007) "Initiating Coordination," *Philosophy of Science*, 74:790-801
Williams, Meredith (1989) *Wittgenstein, Mind and Meaning*, London and New York: Routledge
Wittgenstein, Ludwig (1953) *Philosophical Investigations*, Oxford: Blackwell
Wittgenstein, Ludwig (1981) *Zettel*, Oxford: Blackwell
Zamora-Bonilla, Jesus (1999) "The Elementary Economics of Scientific Consensus," *Theoria*, 14:461-488
Zamora-Bonilla, Jesus (2006) "Rhetoric, Induction, and the Free Speech Dilemma," *Philosophy of Science*, 73:175-93