Original Paper

Bernd Lahno*

Norms as Equilibria

DOI: 10.1515/auk-2016-0121

Abstract: This paper presents a survey on contemporary RC accounts of norms. The characteristic common feature of these accounts is that norms are understood as equilibrium selection devices. The most sophisticated positions driven by this idea are Herbert Gintis' theory of norms as choreographers and Cristina Bicchieri's theory of norms as solutions to mixed motive games. In order to give a comprehensive account of social norms, though, RC theory needs to be substantially extended. In particular, it seems to be impossible in principle to fully understand the concept of normativity and the motivating power of norms within a traditional, pure RC framework.

Keywords: Social norms, rational choice, equilibrium selection, normativity, social preferences

1 Introduction

Social norms are central to our understanding of human interaction. As I understand it the core of a social norm is a social standard that is used as a guide for individual behavior in certain classes of situations.[1] Typically, such a standard may be represented by a rule of the form 'Do x if C applies!' where x defines a way to act and C describes the (possibly complex) condition under which the norm demands that one performs action x. A social norm is said to exist or to be effective to the extent that the corresponding rule is commonly accepted as a standard and in fact guides people's actions.

The theory of Rational Choice (RC) has been one of the most prominent and influential theories of social interaction over the last decades. Although there have been some early attempts within the RC tradition to give an account of social

1 See Lahno 2009, 564f. for a specification of this account.

***Corresponding Author: Bernd Lahno:** Uferstr. 31, 78343 Gaienhofen, Germany, e-mail: b.lahno@me.com

norms[2] these efforts have never been given much prominence. One possible reason for this is probably that social norms may seem to have no peculiar explanatory power from a RC perspective. The core idea of RC is that any action can be understood as instrumentally rational. From this point of view, all a theorist needs to know so as to explain an individual's action is what her aims and her beliefs about the proper ways of realizing these aims are. There seems to be no need to refer to norms in explaining social interaction, once these beliefs and aims are defined.

In recent years, though, RC theorists have started to doubt that the hypothesis of instrumental rationality alone suffices to account for the most important forms of social interaction. And with these doubts a new interest in social norms has arisen within the RC community. The different accounts of norms that ensued all seem to share one basic fundamental idea: that social norms can be represented as equilibrium points. In what follows I will give a critical survey on some of the most important accounts of social norms based on this idea.

The guiding idea of this survey is to introduce those scholars to the field who are interested in a theory of norms but not particularly familiar with the formal peculiarities of RC theory. My aim will be to illustrate the potential as well as the limitations of looking at social norms from a RC point of view without going too deeply into the technical details. It is an impression of the general outlook rather than a sophisticated critical evaluation of the particular representative theories that I seek in this survey.[3]

After a sketch and a short discussion of some of the fundamental obstacles that a RC approach to social norms faces (2), I will introduce the general idea of 'norms as equilibria' using a simple coordination problem as illustration (3). This idea may be extended to interaction with partial conflict as modeled in iterated games representing ongoing interaction (4), and to the variety of actual coordinating mechanisms in social life based on the concept of 'coordinated equilibrium' (5). A first recap (6) records that there are consistent ways to integrate the social fact of norms into a RC theory of social interaction. However, in order to formulate such a consistent theory of norms within an RC framework, it is necessary to extend the framework substantially. One aspect of norm guided behavior, for which the approaches discussed up to *section 5* cannot properly account is social norms' frequent demand for individuals to act against their material interests. The natural way to deal with seemingly 'counter-preferential' choice within the RC tradi-

2 Prominent examples are Opp 1997; 1983; Ullmann Margalit 1977; Coleman 1990; Ostrom 1990.

3 Readers with some familiarity of the general background and an interest in a more detailed discussion of current RC approaches to social norms may refer to Paternotte/Grose 2013.

tion is to introduce so-called 'social preferences'. I will present the model of social preferences suggested by Fehr and Schmidt (1999) as the most prominent representative for this idea today (7). Finally Cristina Bicchieri's theory of norms will be discussed as an attempt to integrate counter preferential choice into a theory of norms as equilibria by introducing preference transformations (8). I will conclude with some tentative remarks about the potential success of a RC approach to social norms (9).

The most important insight from the present survey is that there are consistent ways to account for the norm-guided regularities of social interaction within a RC framework. In order to give a comprehensive account of social norms, though, RC theory needs to be substantially extended. In particular, it seems to be impossible in principle to fully understand the concept of normativity and the motivating power of norms within a traditional, pure RC framework.

2 Obstacles

The traditional reluctance to give a theory of norms much weight within RC is probably grounded on its general difficulties in giving a reasonable account of rule-guided behavior.[4] More particularly, there are three consequences of RC's fundamental axioms that seem to bar the way to a deeper interest in and analysis of social norms.

(1) RC conceptualizes choice and action as 'opportunistic'.

The concept of opportunism is used here in the following sense. Any individual will instantly exploit every opportunity to improve her situation by her individual choices. She will be searching for such opportunities in every single situation and use them immediately, whenever they are detected. What forms an 'opportunity' in this sense is solely determined by the set of actions and the consequences from which an individual can actually choose in a given situation. As such, no motivational force is ascribed to considerations such as 'What will be the overall consequences of my following this route of action in the long run?'.

Opportunism[5] combines extreme agility and responsiveness to changing circumstances with resolute restrictions on the determinants of choice. An op-

4 See Elster 1991 for a classical discussion of the constraints that the concept of instrumental rationality puts on any attempt to account for social norms within a RC framework.

5 Cf. Kliemt 2009, 55 ff for the concept of opportunism as used here.

portunistic decision maker unswervingly takes any opportunity as soon as it occurs. She is never governed by customs or habits. Her behavior may conform with norms (if this suits her interests), but it is not restricted by norms. She does whatever is best, independently of her own past behavior or of what others may think suitable. There is a clear and rigorous restriction on what may influence the decision of an opportunistic decision maker: only the expected future consequences of her own individual choices are relevant for opportunistic choice. This is closely related to the fact, that only specific types of rules are consistent with RC.

(2) The principle of instrumental rationality is fundamental to RC.

According to RC an individual will always maximize utility, i.e. she will choose her actions such that the expected consequences will be optimal according to her preference ordering. This is not to be understood as a claim about the psychological processes in making decisions. As a descriptive theory of action RC does not make a claim about the actual reasoning processes involved in making decisions. It merely postulates a certain factual relation between the beliefs and preferences of an individual and her choices. The principle of instrumental rationality, thus understood, is well consistent with individuals using various sorts of practical rules in making their choices. But, whatever these practical rules prescribe, they must ultimately guide actors to choose those very actions which in fact maximize utility. The important consequence is that only those social norms are comprehensible within RC, which correspond to hypothetical imperatives.

Instrumental rationality, thus, defines a constraint on those social norms that can be viewed as actually effective or as justifiable in RC. The actual strength of this constraint may depend on what is accepted as a proper determinant of an individual's utility function—if regret or a bad conscience are understood as ordinary influences on the evaluation of consequences they will have to be represented in the utility function and thus function as ordinary goals of action. In any case, RC demands that any action can be understood in terms of individual goals and its suitability to realize these goals.

(3) The formal framework of RC is not suited to allow for all types of situational aspects as conditions of action.

According to the principle of instrumental rationality a rational individual will choose that particular option which promises to bring about the best consequences in the light of her individual preferences. A formal representation of a problem of choice will consequently be restricted to those elements that actually have some impact on how consequences relate to action, on what individuals

know about this relation and how they evaluate the potential consequences of action. An important implication from this is that certain aspects of the situation, which might well be important elements in the conditions of a behavioral rule serving as a normative standard, will not be represented in a standard RC model. The model will, for instance, contain no information about previous interactions if these are not reflected in the preferences or beliefs of the (inter-)acting individuals. From a RC point of view such information is, therefore, redundant in a perfectly comprehensive account of individuals' choices. But it may well play an important role within the situational conditions C that form a crucial part of norms. To give but one example, whether a promise was given or not is essential for the demands that the promising norm puts on us. From a rational choice point of view, a promise is just cheap talk unless it does have some actual impact on the potential consequences of action and/or the preferences of the individuals involved.

To summarize: all three fundamental characteristics of a RC theory of social action cited above indicate some sort of tension between RC and a theory of social norms. Opportunism seems incompatible with any commitment to norms. RC appears to be consistent only with norms that can be replaced in principle by hypothetical imperatives. Finally, RC modeling seems too coarse to be capable of capturing all essential elements of social norms.

3 Norms as Equilibria: The Basic Idea

In spite of these obstacles that RC faces, there are also conceptual affinities suggesting that RC might contribute to a better understanding of social norms. One such affinity is that between the concept of a rule and the game theoretic concept of a strategy as developed and defined in RC. A behavioral rule defines and prescribes a certain way to act in any situation of a certain kind. Compare this to the definition of a strategy in game theory: given a game Φ, a strategy of a player i in Φ is a function that assigns an action of i to any situation in the course of the game that calls for a choice by i. In short, a strategy may be understood as a general rule that prescribes what to do in any contingency that may arise in a situation represented by the game. Now, a norm tells us what the 'right' way to act under certain conditions is. This corresponds to what RC defines as the 'rational' way of acting. A norm for some situation S can thus be represented by an optimal strategy in a corresponding game Φ that represents S. In a standard strategic game optimal play is usually represented by equilibrium strategies. It seems very plausible, then, to identify norms with equilibrium strategies or (if the norm makes

a claim on the behavior of every participating individual of some situation) with equilibrium points.

The idea that norms are equilibria or equilibrium strategies has been put forward by many scholars within the RC tradition.[6] The core idea and some of its more important implications can be illustrated by discussing a simple game of coordination as given in *figure 1*.

		Berta: **Wrestling**	Berta: **Movie**
Adam	**Wrestling**	**2 | 1**	**0 | 0**
	Movie	**0 | 0**	**1 | 2**

Fig. 1. Battle of the Sexes

This game is known as the 'Battle of the Sexes' (BoS). One might think of two individuals Adam and Berta deliberating about where to go in the evening. Each can either go to the cinema and watch a movie or attend a wrestling match at the local gym. Both would prefer to spend the evening together but they do not have the chance to arrange where to meet. Adam, the row player, prefers to meet Berta at the gym (represented by a payoff 2),[7] but if Berta goes to the movies he would rather go there (payoff 1) than watch the wrestling match alone (payoff 0). Berta, the column player, prefers to meet Adam at the movies (payoff 2). But like Adam she prefers any state in which she meets Adam to spending the night on her own (payoff 0).

Look at the situation from Adam's point of view. What Adam should do depends on what he thinks Berta will do. If he thinks that Berta will go to the movies, he should do the same, and if he thinks that she will go to the gym, then he should also go there. He prefers meeting at the gym but that does not help him make his choice because Berta is facing the same problem, except that she would prefer to

6 Proponents of such an account are numerous. Among the more prominent predecessors of those scholars cited below are Thomas Schelling 1960; David Lewis 1969; Robert Sugden 1986; Michael Taylor 1987; and Ken Binmore 2005.

7 In each cell the first number represents the payoff of the row player and the second the respective payoff of the column player.

meet at the cinema. If both of them go for a meeting at their respective favorite place, they will not meet.

If both choose Wrestling, then none will have a reason for regretting his or her choice. Given the choice of the other each does as best as he or she can. Or, in the language of game theory: the strategy profile (Wrestling, Wrestling) is a Nash-equilibrium. But, of course, there is a second Nash-equilibrium, namely (Movie, Movie). It seems completely unclear how and on what pattern of behavior the two can coordinate. A behavioral standard solves this coordination problem. For, if it is common knowledge among the players that the default meeting point for such situations is at the wrestling match, both would certainly go to the local gym. The same would be true, if the standard said, for instance, that one should coordinate on what the man prefers.

This is the basic idea of 'norms as equilibria': A norm (or a system of norms) determines a pattern of 'right' behavior for certain situations, a definitive way to act for each of the participating individuals in such a situation. Norms can thus be represented by a strategy profile of the game that represents the situation. The corresponding behavioral patterns may prevail only if (on the whole) no individual has an incentive to deviate from the pattern as long as (almost) all others comply with it, i.e. if the corresponding strategy profile is a Nash-equilibrium. So any social norm that is actually effective (or could become effective in principle) in some sort of situation can be represented by an equilibrium point of the game representing the sort of situation.

Note that it is not claimed that any Nash-equilibrium in a game representing the abstract structure of some sort of situation actually or potentially constitutes a norm. Rather, the theory of norms as equilibria defines only a necessary condition for social norms to be effective and produce stable behavioral patterns. The theory does not tell us which equilibria will actually materialize as social norms. As a matter of fact, social norms appear as elements external to RC. They are added to standard RC in order to tackle a problem that cannot be solved adequately within RC, namely the problem of equilibrium selection in cases of multiple equilibria. RC tells us in part, what norms must be like to make us capable of solving such problems, but it does not tell us very much besides this: that is, it does not determine which particular norms may become effective and what the processes are that render these norms effective.

4 Cooperative Norms

Our simple example uncovers a basic fact about norms: norms do not have to be fair. They can produce stable behavioral patterns and expectations even though these patterns clearly privilege a certain subgroup of the individuals involved and even though all individuals, including those disadvantaged, are perfectly aware of this. However, the conflict of interest incorporated in a simple coordination game such as BoS is comparably low. In contrast, many norms seem to regulate behavior in situations characterized by more severe conflicts of interest. The 'Prisoners Dilemma' (PD) is commonly used to illustrate a situation that displays such conflict and, therefore, some demand for cooperative regulation.

If we look at the isolated single game as given in *figure 2*, the theory of norms as equilibria does not seem to provide a solution to the problem of cooperation. There is only one equilibrium point in the PD, namely (D, D), mutual defection. No norm seems necessary to motivate defection in the PD. D is actually a dominant strategy for both players: it is each player's best answer to any choice by the other. A cooperative behavioral pattern seems to be unattainable although it would serve the interest of both individuals.

	C	D
C	3 | 3	0 | 5
D	5 | 0	1 | 1

Fig. 2. Prisoner's Dilemma

The situation may change if a whole series of such games is played instead of just one isolated PD. In the case of iterated games, players may make their choices dependent on what they observed in the past which, in turn, may produce cooperative incentives for both. A standard way to model such iterated play is a so-called supergame. In a PD supergame two players A and B play a standard PD (e.g. as in *figure 2*) with constant payoffs at every point in time of a series t= 0, 1, 2, … until the series terminates. After every round a random mechanism stops the series with a constant probability 1–ρ. The payoff that players earn is given by the sum of their payoffs in all stage games before the series is terminated. These rules including the value of ρ are common knowledge among the players.

There is an infinite number of strategies in a PD supergame, among them conditionally cooperative strategies such as:

TIT FOR TAT: 'Choose C at t=0; for any t>0 choose C if your partner chose C at t–1, and D if your partner chose D at t–1.'

TRIGGER: 'Choose C at t=0; for any t>0 choose C if your partner chose C at all preceding points in time t'<t, otherwise, choose D.

There are of course also non-cooperative strategies such as ALL D (defect in any round irrespective of what your partner does or did), and an uncountable number of strategies that prescribe more or less sophisticated ways of reacting to certain patterns of play.

TIT FOR TAT became quite popular in the 1980s when it succeeded in two computer tournaments with iterated PDs that were organized by the political scientist Robert Axelrod (1984). (TIT FOR TAT, TIT FOR TAT) is an equilibrium in the PD supergame if the probability 1–ρ that the series terminates at any point in time t is sufficiently low. Axelrod and others concluded, therefore, that stable behavioral patterns may arise which correspond to conditional cooperative strategies in the PD.

The situation in the iterated PD is, in fact, much like the situation in the Battle of the Sexes game discussed above. In contrast to the single PD, which has one single equilibrium and a clear solution in (D, D), there are (depending on the 'discount parameter' ρ) many Nash-equilibria in the PD supergame[8], among them many which are cooperative or partly cooperative. Any equilibrium—as any strategy profile more generally—can be understood as a more or less complex system of rules to guide the behavior of all individuals involved in all situations they might face in the course of a PD supergame. RC tells us that only equilibria can form stable patterns of behavior. If a social norm or a system of social norms produces a stable cooperative pattern of behavior in ongoing social interaction of the PD form, the norms or the system of norms must, therefore, correspond to an equilibrium point of the iterated PD.

In contrast to BoS, the PD—and with it also the PD supergame—displays much conflict of interest. However, just as in BoS, there is an equilibrium selection problem in the PD supergame. And just as in BoS, there is no canonical solution to this problem within RC. A social norm such as 'cooperate with those that are known to you as cooperators!' may solve this problem by defining a stable equilibrium for ongoing interaction.

8 There is generally an uncountably infinite number of equilibria in any non-trivial super game as the so-called folk theorem tells us (see e.g. Fudenberg/Tirole 1991, 150–160, for an overview and more references regarding the 'folk theorem').

5 A Refinement: Correlated Equilibrium

The norm (1) 'Do your part in the behavioral scheme that the man prefers!' which was introduced above as a solution of the BoS problem exemplifies a property that many social norms possess if looked at from a RC perspective. The norm refers to an aspect of the world that is not reflected in the game theoretic model of the situation, namely the sex of the players. One can imagine many rules that would solve the equilibrium selection problem of the Battle of the Sexes by reference to some event or state of affairs which is not reflected in the game. For instance, the rule could make the decision dependent on the day of the week: (2) 'Meet at the gym on ordinary weekdays and in the cinema at the weekend!'. Or it could make its recommendation dependent on past behavior: (3) 'If you met at the gym last time, go to the cinema, if you met at the cinema, go to the gym!'. The concept of 'correlated equilibrium' seeks to account for the fact that decisions might be coordinated by making them dependent on some external event. A correct formal definition of the concept would exceed the scope of this survey. However, those aspects that may be relevant in a theory of social norms may be elucidated by the metaphor of a 'choreographer' as introduced by Herbert Gintis (2009; 2010).

Suppose a choreographer observes some external aspect of the world. Depending on what he observes, he gives compound advice for action by suggesting a strategy to each individual player who finds himself in some sort of situation represented by a game. Roughly, this conditional advice produces a correlated equilibrium iff for each individual, following the advice is a best response to all other players following the advice, given what the individual knows about how the choreographer determines his advice (i.e. how it depends on his observation). A social norm, then, can be understood as a choreographer telling us what to do in certain kinds of situations. The norm will be stable if the compound advice of this choreographer forms a coordinated equilibrium in these situations.

Norm (1), saying that coordination is to be based on what the man prefers, corresponds to a choreographer telling both players to go to the wrestling show irrespective of the state of the world in the BoS. Norm (2) represents a choreographer who looks at the calendar and advises the players to choose Wrestling during the week and Movies at the weekend.

So far, in all our examples of (potential) norms, following the norm results in realizing a Nash-Equilibrium of the underlying game. However, a coordinated equilibrium is not necessarily a Nash Equilibrium of the original game. The theory of norms as choreographers suggesting action according to some coordinated equilibrium is, therefore, a proper extension of the theory of norms as Nash-

equlibria. Another simple example may illustrate this. Consider the game in *figure 3*.

The game represents the abstract structure of two individuals simultaneously reaching a road crossing: one heading north on the first road and the other heading west on the second crossing road. If both continue to drive, they will collide ((0, 0)). If both stop in order to find an understanding on who is to pass first, both will lose time ((1, 1)). If only one of them stops and the other crosses the junction, only the one stopping will lose time ((1, 2) or (2, 1)).

	Stop	**Go**
Stop	**1 | 1**	**1 | 2**
Go	**2 | 1**	**0 | 0**

Fig. 3. Traffic Game

The game has two Nash-equilibria in pure strategies[9] namely (Stop, Go) and (Go, Stop). But it is unclear how players would coordinate so as to choose one of them. Traffic rules such as 'Drivers on streets connecting North and South have priority over drivers on roads in east-westerly directions' or 'Give way to vehicles coming from the right' solve the problem. Both can be represented by a choreographer who gives his advice dependent on some originally irrelevant aspect of the situation. So far, then, both recommend equilibrium play in the given traffic game. But now consider the following solution:

A set of traffic lights is installed. There are three possible states of the system: RG, GR and RR, where the first character indicates the state of the traffic lights on the North-South street—R indicating red (Stop), G indicating green (Go)—and the second character represents the state of the traffic lights on the East-West road. States successively change from RG to RR to GR to RR to RG and so on. Obviously, RR does not recommend equilibrium play in the traffic game. But it still makes

9 As in the examples above we ignore equilibria in mixed strategies to keep the analysis simple. Mixed strategies are probability distributions on pure strategies. It seems unlikely that they represent the behavioral standards of social norms.

sense, as it reflects the fact that drivers may not be able to ground to an instant halt when their light changes from G to R.

Notice that the traffic light system gives different private signals to the drivers. When the driver A on the North-South street sees a green light, he knows that a driver B on the crossing East-West road is shown a red light and will stop as long as he complies with the choreographer's suggestion. So following the choreographer is also optimal for A. However, if his traffic lights show R, then A only knows that B's traffic lights might show either R or G. Let ω_1 = prob(RG) be the probability that the state for the system is RG, and, similarly, ω_2 = prob(RR). The probability of RR when A is shown a red light is $\omega_2/(\omega_1+\omega_2)$. Given that B complies with the rule, it is best for A to do the same, iff his payoff for following the rule and choosing Stop (which is 1 irrespective of the state of the system) is larger then his expected payoff for deviating and choosing Go (2 in case of RR, 0 in case of RG):

$$1 \geq \frac{\omega_2}{\omega_1+\omega_2}\cdot 2+\frac{\omega_1}{\omega_1+\omega_2}\cdot 0 \Leftrightarrow \omega_1 \geq \omega_2$$

B's situation is symmetric. So, if the time span of RR is sufficiently small as compared to red-green phases the traffic light system and the norm 'Stop at Red, go at Green' will constitute a 'coordinated equilibrium' although it will not in general suggest equilibrium play in the original Traffic Game.

The theory of norms as (constituting) coordinated equilibria is, therefore, a genuine extension of the theory of norms as equilibria. It explains a wider class of behavioral regularities as the outcome of a wider class of social norms. Gintis points out another attraction of his coordinated equilibrium approach to norms: the assumption that (Bayesian) rational individuals will choose their strategies according to a coordinated equilibrium can be based on much less demanding epistemic premises than the corresponding assumptions in the case of Nash-equilibria. An attempt to explain and explore this difference requires considerable technical effort,[10] which would far exceed the scope of this overview. Gintis' observation, though, reminds us that there are epistemic preconditions for the claim that rational individuals comply with a Nash- or correlated equilibrium: minimally, individuals must share their conjectures about other individuals' choices (because individual rationality will guide them to do their part of the equilibrium only on the basis of these beliefs). The theory takes this as given, it is part of the explanans, not the explanandum. In the terminology of norms: the RC

10 A clarification would presuppose an introduction to the theory of epistemic games. See Gintis 2010, 256ff.; 2009 chapters 4, 7.

theory of norms explains why people would follow a norm *on the basis that* they have a commonly shared expectation that (sufficiently many) others comply with the norm. It sets necessary conditions on what the content of a norm may be and on what beliefs can be expected among rational players, but it does not itself explain, why individuals would have these expectations. However, the expectations of the norm addressees seem to form an essential part of norms as a social fact. There is a desideratum in the RC theory of norms at this point. In order to solve the puzzle why particular norms evolve while others do not the theory needs the input of sociological or psychological theories.[11]

6 A First Recap

Let us briefly suspend the survey for a first, short recapitulation. An account of social norms as equilibrium selection devices or as choreographers that guide individuals to form a correlated equilibrium elucidates essential characteristics of social norms. Norms are fundamentally based on mutual expectations of norm-conforming behavior. They may be part of complex systems of norms that include some rules of sanctioning. Norms are not necessarily fair, and the behavioral pattern they induce may not be efficient. Norms may extend the range of choice determinants, which means that "life based on social norms can be significantly qualitatively richer than the simple underlying games that they choreograph" (Gintis 2010, 255).

Contrary to what our considerations about the obstacles faced by RC may have suggested, there is thus a consistent way to integrate a conception of social norms into RC. On this conception, following a norm is utility maximizing and, therefore, consistent with instrumental rationality. This does not imply that actors must actually follow the norm for the reason that this maximizes their utility. RC does not make a psychological claim about the motivation of individuals. So, the theory is also consistent with the assumption that (at least some) actors conceive of the norms as categorically binding and feel committed to the norm irrespective of their individual inclinations. Norms can be genuine, subjective reasons for action. What RC excludes, though, is that such reasoning may prevail if the norm (continuously) prescribes suboptimal behavior.

An assumption that norms have some independent motivational power may, in fact, fill a gap in the argument that norms serve as equilibrium devices. To elaborate, rational individuals in the strict sense of RC have a reason to follow a norm

11 Gintis (2009, 143, 162) explicitly emphasizes this point.

only if they expect others to do so as well. At the same time, they will be aware of the fact that others are in exactly the same position: Everybody has a reason to follow the norm only if others can also be expected to observe this norm. Accordingly, this impasse can be overcome only if there is a reason (at least for some) to follow the norm that is independent from expectations about others' behavior[12]. Now, if it were commonly known that some (sufficiently many) individuals are categorically motivated to act as the norm prescribes, this would provide such a reason. A similar argument applies to the case of norms as choreographers that implement a coordinated equilibrium. Appealing to a basic theorem of Aumann (1987), Gintis argues that Bayesian players who choose a coordinated equilibrium do so because they share a common expectation (a so called common prior) about the likelihood of the corresponding behavioral pattern. He then goes on to state that this theory does not provide any reason "as to why players would have common priors and why they would choose any particular equilibrium over any other" (Gintis 2009, 8). A normative, motivating power of social norms—although itself something in need of explanation—would provide such a reason, whether directly by guiding action or indirectly by guiding expectations. But such reasons cannot be conceptualized within traditional RC. They are external and have to be added to RC in order to form a comprehensive account of social norms.[13]

The two variants of a theory of norms as equilibria introduced so far cannot by themselves give a comprehensive account of the phenomenon of social norms. Whilst they elucidate some essential aspects of social norms, they fail to account for others. As we have just pointed out, the normative power of norms is alien to the RC framework. Norms define duties and rights, but duties and rights do not fit well with the consequentialist framework of RC. Neither are they appropriately captured by the rules of the game that reflect the external constraints in RC models of social interaction, nor can they be adequately depicted by individuals' preferences over consequences of action or their beliefs. Moreover, norms are associated with normative expectations, but the beliefs of RC models of interaction are entirely cognitive: their content is of the form 'A is or will (probably) be the case'. RC

12 This is actually a general argument showing that instrumental rationality in the sense of RC alone does not offer a reason to choose one equilibrium strategy rather than another. See Lahno 2007 for a more detailed exposition of the argument and further references.

13 As a matter of fact, Gintis introduces the concept of an 'α-normative predisposition' to represent the (constrained) motivating power of a norm. But this concept is inconsistent with the standard assumptions of instrumental rationality. Moreover, Gintis uses the concept in an attempt to explain counter-preferential choice in cases of incomplete information. He does not refer to it to explain the choice of equilibrium strategies.

simply does not provide a conceptual framework to account for contents of the form 'A should be the case'.

There are other aspects of social norms that do not seem well captured in the theories introduced so far. Firstly, theses theories are essentially static and, therefore, do not account for the dynamic character of social norms as, for instance, their tendency to adapt to changing social circumstances or to be transferred to different contexts that share certain features with the original context. Secondly, the theories inform us about the constraints on what kind of rules or systems of rules may be effective under what conditions, but they do not elucidate how rules become effective. In the same vein, the theories specify the conditions that cause a social norm to become ineffective, but they do not tell us anything about the process of the breakdown. Thirdly, from an RC point of view, one would expect that a behavioral pattern immediately breaks down as soon as (sufficiently many) individuals become aware of opportunities to gain by deviating behavior. However, empirical evidence shows that norm compliance is much more stable than this would suggest. Finally, there is another general problem of an RC theory of social norms that is related to its fundamental difficulties with normativity: while RC rests on the principle that choices are generally directed at maximizing utility, many norms seem to demand counter-preferential choice. To this problem we turn in the next section.

7 Social Preferences

Strictly speaking, there is no counter-preferential choice within a traditional RC framework. However, the theory of revealed preference is not committed to any specific substantial criteria of preference.[14] A rational individual A in the sense of RC may well prefer a consequence K_1 to a consequence K_2, although K_1 is associated with some cost in terms of A's material interests as long as he does so in a consistent way. To account for such choices, a RC theorist will typically distinguish between payoffs which represent material gain and effective utilities which represent choice as actually revealed in the behavior of a rational agent. Hence, the effective preferences of an altruist A interacting with individual B are often assumed to be represented by a convex combination $u_A=\alpha x_A + (1–\alpha)x_B$ where x_A and x_B represent the payoffs of A and B. A similar idea of 'social preferences' may be used to account for norm conforming behavior, such as the compliance with

14 An accessible (and critical) introduction to the fundamental ideas of the theory of revealed preference is Sen 1973.

fairness norms. A prominent example is the analysis of the behavior observed in a so-called ultimatum game (Güth et al. 1982).

Suppose a certain amount of money X is to be divided among two actors A (the 'proposer') and B (the 'responder'). A first has to declare how much he is willing to concede to B. If B accepts A's suggestion the money is divided accordingly. If B rejects the suggestion, neither of them receives anything. *Figure 4* illustrates such a situation, assuming that X = 3 and that only whole units of money can be handed out.

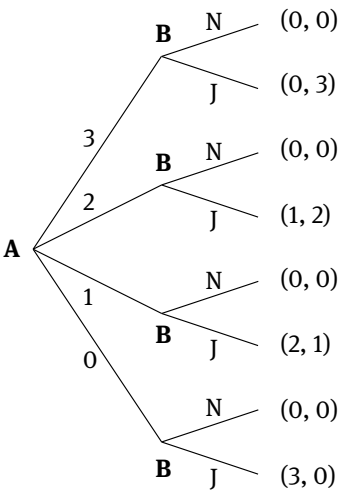

Fig. 4. Ultimatum Game

If both actors care for their own payoffs only, a rational responder B would accept any positive offer of proposer A. Consequently, RC would predict that A offers the minimal amount and B accepts.[15] This prediction has been tested many times with different variants of the game and in various cultural contexts.[16] As one would, of

15 In *figure 4* the first number in any terminal node represents the payoffs of the first player A, and the second number the payoff of B. The game has two subgame perfect equilibria, namely (s_A, s_B) and (s_A^*, s_B^*) with s_A denoting A's strategy to offer 0, s_B denoting B's strategy to accept any offer, and s_A^* denoting A's strategy to offer 1, s_B^* denoting B's strategy to accept any strictly positive offer. If any real-valued amount can be offered the unique subgame perfect equilibrium and the solution by backward induction is (s_A, s_B).

16 See Henrich et al. 2004 for an overview.

course, expect, this prediction cannot be verified: Rather, it has been found that as a rule proposers offer between 40% and 50% of the amount, which responders generally accept, while they reject offers which are substantially smaller.

These results cannot be explained by altruism represented by a social utility function as mentioned above[17]. Another, more plausible attempt to explain the results is by reference to the agents' fairness concerns. There are several attempts to integrate fairness concerns in game theoretical models. The most prominent among them is probably the approach by economists Ernst Fehr and Klaus M. Schmidt (1999)[18]. Fehr and Schmidt argue that a concern for fairness can be incorporated in the utility function of an individual by subtracting a certain value according to the inequality observed. Suppose the outcome of an N-person game is given by $x = (x_1,x_2,..x_N)$ with x_i $(1 \leq i \leq N)$ denoting the payoff of individual i. Then, according to the Fehr-Schmidt model, an individual i evaluates outcome x according to the following utility function:

$$u_i(x) = x_i - \alpha_i \frac{1}{N-1}\sum_{j \neq i} \max\left\{x_i - x_j; 0\right\} - \beta_i \frac{1}{N-1}\sum_{j \neq i} \max\left\{x_j - x_i; 0\right\} \quad (\text{with } \alpha_i \geq \beta_i \text{ and } 0 \leq \beta_i < 1)$$

α_i denotes the weight that i attaches to (the average) inequality in favor of himself, β_i accounts for the inequality that favors others. In a simple 2 person game as in *figure 4* the equation reduces to:

$$u_i(x) = \begin{cases} x_i - \alpha_i(x_i - x_j) & \text{if } x_i \geq x_j \\ x_i - \beta_i(x_j - x_i) & \text{if } x_i < x_j \end{cases}$$

As a matter of fact, this technically quite simple model is consistent with many empirical observations, in particular with evidence from experiments with ultimatum games. Fehr and Schmidt emphasize that it also explains the different consequences of fairness concerns in varying contexts, such as differences between

17 An altruist responder would accept any positive amount while an altruist proposer knowing this would offer either a maximum or a minimum amount depending on whether α is smaller or larger than 0.5.

18 Another similar but axiomatically founded approach is Bolton/Ockenfels 2000. Rabin 1993 introduced a widely cited, more complex game theoretic model of fairness using 'psychological game theory'. His approach is technically most demanding, whilst its applicability to real life social interactions is restricted.

bilateral bargaining and competitive markets.[19] However, as they also note, the model is not sensitive to all context differences.

Assume that the money which A is to divide between himself and B is a bonus payment for the special effort that A invested in a common project. Compare this to the case that it is B's bonus payment and A is commissioned to hand it over to B after taking some of the money as a compensation for this service. These are severe contextual differences and one would, of course, expect them to be reflected in corresponding behavioral differences. But the abstract structure of the situation as given by the extensive form game (as in *figure 4*) remains the same. Consequently, if there is a difference in outcomes although the payoffs—i.e. the amount to be divided and the rules of the game—remain the same, it has to be reflected in different parameters α_i and β_i.

From this point of view, then, a fairness norm is a function assigning pairs of parameters α_i and β_i to specific kinds of situations such that each of these pairs defines a specific utility transformation as given in the equation above. Note that this function is an external element in any RC model of fairness-driven behavior. The situational differences as illustrated in our example above cannot be accounted for within a RC model, except by modifying the preferences of the actors. Hence, functional dependency is included in the rules of the game; it is taken as given and cannot be explained by the theory.

Note also that this approach to (fairness) norms remains entirely within the consequentialist framework of RC. In contrast to the theory of norms as equilibrium selection devices, which transcends standard RC by adding decision making rules in cases of equilibrium selection problems, utility maximizing is the only effective decision making rule within this approach. From this point of view, a norm tells us primarily how we are to evaluate the possible consequences of action, and only indirectly—by an application of the maximizing utility rule to the transformed model of the situation—how we should act. Norms appear to be standards of evaluation rather than standards of conduct (cf. Lahno 2010).

While the theory of norms as equilibrium selection devices refers to norms as a way to account for behavioral regularities that cannot be explained by reference to instrumental rationality in the strict sense of RC alone, the theory of Fehr and Schmidt introduces norms as a mechanism that motivates us to act (occasionally) against our material interests. Obviously social norms do both: Some

19 The original ultimatum game is understood as representing bilateral bargaining. The context of a competitive market may be modelled by versions of the game in which several proposers compete for an interaction with the responder or in which several responders compete. In both cases, the model predicts that fairness concerns are not effective (in the sense of not affecting behavior), which is consistent with what is actually observed.

of them sometimes tell us what to do if our goals alone do not suffice as a guide to action, and some of them sometimes tell us what our goals should be. This suggests a combination of the two approaches. In the last section of this survey, we will shortly discuss such a combined approach.

8 A Comprehensive Attempt

Cristina Bicchieri's theory of norms (2006) is probably the most comprehensive attempt to give an account of norms within the Rational Choice tradition. We can provide only a sketch of her key ideas here.

Bicchieri discriminates between social norms on the one hand and conventions and descriptive norms on the other. She argues that conventions and descriptive norms are correctly described as equilibrium selection mechanisms. Once an individual identifies a situation as one in which a conventional rule or descriptive norm applies, it typically suffices for her to believe that sufficiently many other individuals will conform to the respective convention or descriptive norm. If so, she will also be motivated to comply with this convention. However, a more complex motivation is required in order to motivate compliance with a social norm. The reason is that social norms regulate behavior in situations that are characterized by some amount of conflict, i.e. situations that are—in the language of game theory—so-called mixed motive games such as the Prisoners' Dilemma or the Ultimatum Game. A social norm will typically proscribe actions that go against the interests of the individuals as given independently of the norm. In other words, a social norm typically demands behavior that does not lead to equilibria in terms of material interests. Accordingly, to become effective, a social norm must be backed by additional motivations. Bicchieri argues that these additional motivations are conditional in two respects.[20]

First, a situation must be acknowledged as one in which the social norm applies. Norms are local in the sense that their key concepts such as 'fairness', 'reciprocity' or 'trust' are differently interpreted depending on the specific context, and thus "the expectations and prescriptions that surround them vary with the objects, people and situations to which they apply" (2006, 76). The general idea is then that a situation may offer situational cues which—if the situation is appropriately framed—might become salient to the individuals and trigger the realization that the norm is applicable in a given situation. Once a situation is categorized

20 This constitutes an essential difference from genuine 'moral norms' which motivate unconditionally. See Bicchieri 2006, 20f.

as being of a certain type, a 'script' is activated that describes a stylized sequence of behavior which is appropriate in this context and defines actors and interlocking roles. This script thus potentially supplies a shared understanding of what is supposed to happen and what one and others ought to do.

This is where the second condition of norm compliance occurs. The script will become motivationally effective only if individuals believe that sufficiently many others will actually expect them to conform and, possibly, that they are willing to sanction non-compliance. Hence, there are two types of expectations that underlie norm compliance:[21]

(a) *Empirical expectations:* individuals believe that a sufficiently large subset of the relevant group/population conforms to the norm in situations of type S and either

(b) *Normative expectations:* individuals believe that a sufficiently large subset of the relevant group/population expects them to conform to the norm in situations of type S;
or

(b') *Normative expectations with sanctions:* individuals believe that a sufficiently large subset of the relevant group/population expects them to conform to the norm in situations of type S, prefers them to conform and may sanction non-compliance.

Empirical expectations along the lines of (a) suffice in motivating individuals to conform with a convention or descriptive norm, because the underlying decision problem is typically a coordination problem. In contrast, social norms address problems of cooperation and, therefore, must serve to overcome a partial conflict of interest between individuals. They do so by motivating individuals to comply with the normative demand even if this goes against their individual interests as given independently of the norm. Bicchieri assumes that the fact that sufficiently many others expect that an individual will comply actually produces a motive strong enough to make her comply. This is what normative expectations according to (b) or (b') ensure.

If a social norm becomes effective because a situation is commonly acknowledged as covered by the shared norm, and individuals' expectations are in conformity with (a) and (b) or (a) and (b'), then the norm performs two tasks. Firstly, it produces a new motive within individuals: Individuals want to conform to the demands of the norm. Bicchieri mentions several potential reasons why this may

21 Cited from Bicchieri/Muldoon 2014, see also Bicchieri 2006, 11.

be so: an individual may fear "resentment and unpleasant consequences for the transgressor", she may "desire to please others by doing something others expect and prefer one to do", or she simply "accepts others' normative expectations as well founded" (all citations: Bicchieri 2006, 23). Secondly, the norm tells individuals what to do in light of such motives. The first effect corresponds to a preference transformation within an RC framework similar to the one discussed above in the context of fairness. The second can be described as a (set of) behavioral rule(s) which should define an equilibrium point in the transformed game from an RC point of view.

The example of a Prisoners' Dilemma as in *figure 3* helps to illustrate the main ideas. A cooperative norm would become effective if an individual has beliefs according to (a) and (b) or (a) and (b'), i.e. if she believes that sufficiently many others will cooperate and expect her to cooperate (in the case of (b'), she will additionally expect that others may sanction uncooperative behavior). Having such expectations will make the individual prefer to conform with the cooperative norm. Bicchieri proposes an explicit model for the respective preference transformation. This particular model seems tentative and not essential for the general idea. Nevertheless, it offers a good way to illustrate the basic points. Assume that the norm of cooperation demands that A cooperates if B will cooperate and vice versa. If s is a strategy profile reflecting the violation of a norm by a singular actor, then the utility of s for an actor i is his payoff diminished by an amount proportional to the maximum loss caused to one of the individuals involved. In our PD game, this amounts to:

$u_A(C, C) = x_A(C, C)$;
$u_A(C, D) = x_A(C, D) - k_A \cdot (x_A(C, C) - x_A(C, D))$;
$u_A(D, C) = x_A(D, C) - k_A \cdot (x_B(C, C) - x_B(D, C))$;
$u_A(D, D) = x_A(D, D)$.

where x_A denotes the payoff, u_A the (transformed) utility of A and k_A the (constant) weight that A gives to a loss due to a norm transgression. Analogue equations apply to B. If we assume that $k_A = k_B = 1$ we get the following transformation of the original PD in table 3 if both actors classify the situation as one in which the cooperative norm applies and both have the necessary expectations according to (a) and (b) or (b') (*figure 5*):

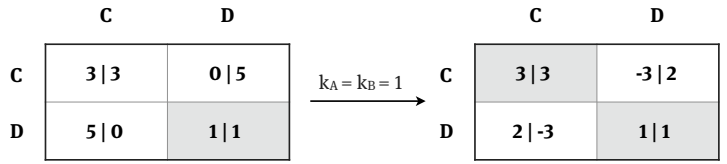

Fig. 5. Preference transformation by a cooperative norm

The resulting game has two equilibria, (C, C) and (D, D), and, thus, poses a typical equilibrium selection problem.[22] In addition to initiating the preference transformation the social norm serves as a coordination device telling the individuals to do their part in the cooperative pattern (C, C): the norm induces cooperation which is an equilibrium in the transformed, but not in the original game.

Bicchieri's model addresses some of the main shortcomings of RC theories of norms as introduced above. Its main achievements are:

(1) In contrast to standard theories of norms as equilibria, Bicchieri's model gives an account of norms' motivating power that transcends individual interest.
(2) In contrast to theories of social preferences, Bicchieri's model explicitly accounts for the context dependency of social norms' efficacy.

The model thus combines the strengths of the traditional RC theories whilst, by specifying the contextual dependency, goes beyond these accounts. Notice that it does so without abandoning the framework of instrumental rationality that is held to be fundamental and characteristic of RC. As Bicchieri is eager to stress, the motivational power of norms becomes effective only indirectly; "[t]he direct underlying motives are the beliefs and desires that support the norm" (Bicchieri 2006, 22). Importantly, these beliefs and desires unfold their motivating power in exactly the way that RC demands: They make people do what is best, given their beliefs and desires.

Given the emphasis that Bicchieri places on the motivating power of norms and to normative expectations, one might expect that she introduces a non-instrumental ought into RC. But her firm commitment to the framework of instrumental rationality prevents her from doing so. This becomes particularly clear when we take a closer look at what Bicchieri calls the condition of 'normative

22 The game is not a game of pure coordination in the strict sense of Lewis 1969. But the coordination problem is serious, because a pareto-efficient equilibrium (C, C) competes with a risk dominant equilibrium (D, D). A game of this form is commonly known as a stag hunt game.

expectations' in (b) or (b'). As she emphasizes, condition (a) may refer to 'empirical' as well as 'normative' expectations in the sense of expressing the belief that others 'ought' to conform to the norm (2006, 14). But, even then, this 'ought' may be a prudential 'ought' without any sense of obligation. Something very similar applies to condition (b):

> "[…] not only do I expect others to conform, but I also believe they expect me to conform. What sort of belief is this? On the one hand, it might just be an empirical belief. If I have consistently followed R in Situations of type S in the past, people may reasonably infer that, ceteris paribus, I will do the same in the future, and that is what I believe. On the other hand, it might be a normative belief: I believe a sufficiently large number of people think that I have an obligation to conform to R in the appropriate circumstances." (2006, 15)[23]

So the condition of 'normative expectations' may refer to normative expectations, but it may also refer to purely empirical expectations. However, it is crucial to note that even those beliefs which Bicchieri labels 'normative' are, in fact, empirical beliefs: They are beliefs about a certain empirical fact, namely the fact that *others* entertain certain normative judgments. Accordingly, there is nothing normative about Bicchieri's 'normative' beliefs. Rather, they are empirical expectations about the normative judgements or empirical expectations of others.

All this is quite understandable, given Bicchieri's commitment to instrumental rationality. 'Normative' expectations may play a role in developing the determinants of choice in case of norm-guided behavior, but they are not part of these determinants. They cannot be, because they are not the kind of things that serve as determinants of choice within traditional RC: (cognitive) beliefs and preferences.

A similar statement applies to obligation as a motive to conform with a norm. I may conform because I am obligated or because I believe that I am obligated. But this is only one of various ways in which my preference for norm-compliance may be formed. Strictly speaking, the preference ultimately guiding my decisions is not a preference for norm-compliance but a preference for the corresponding consequences of action which somehow encapsulates the compliance-preference. But, of course, a preference for the consequences of norm-conformity is by no means equivalent to an obligation or the belief that there is such an obligation. I can be obligated or believe to be obligated to conform with the norm without actually having a corresponding preference for the consequences and vice versa.

As Bicchieri rightly points out, the motives to follow a norm may be of various kinds. I may conform because I fear the sanctions of others, or simply because of habit, because I want to meet the expectations of others, because I think that oth-

23 In later publications the stress is exclusively on (empirically) expecting normative expectations of others; see e.g. Bicchieri/Erte 2009; Bicchieri 2010.

ers have a justified claim on me or I may just feel obligated, whatever others think. Bicchieri's theory can account for all these cases by representing these motives in the preferences for the potential consequences of action. But this implies that the theory no longer discriminates between the different motives. On the level of RC theory, they appear as indistinguishable forms of preference transformation.

Just as context sensitivity and the conditions of norm activation, normativity is something external to RC. True, Bicchieri's theory can integrate these elements of a theory of norms by displaying their consequences for individual preferences. However, the RC framework does not provide for any motivating power beyond the means-end scheme of instrumental rationality. Consequently, it remains blind with regard to the specific motives stemming from obligation. Just as with any other genuine Rational Choice theory of norms, Bicchieri's theory has to be informed about the motivational power of norms by some theory that transcends RC.

9 Conclusion

Our observation in the first recap above remain true. An account of social norms can be consistently integrated into a RC framework. This, however, presupposes that the RC perspective on social interaction is essentially expanded. As the theories introduced above, Bicchieri's theory needs some theoretical and empirical input form the behavioral sciences and psychology to give an RC account of norms. This is true not only for the motivational power of norms but also for their context sensitivity, the mechanisms of norm activation, and, most fundamentally, for the decision making rule and the content of the norm itself. All these aspects can be somehow represented in the theory and thus serve to explain social behavior. But none of these aspects can itself be entirely comprehended or explained within a pure RC framework.

This does not imply that RC merely provides an alternative way of presenting an independently given theory of norms without offering any substantially new insights. At least two such insights contribute significantly to a theory of norms. Firstly, the RC framework sets constraints on the patterns of behavior that may be induced by social norms. Secondly, once a social norm is represented in the framework (e.g. by a suitable combination of preference transformations and behavioral rules), it may produce substantial predictions about social behavior under different circumstances and in response to circumstantial changes. In fact, Bicchieri derives a number of empirical hypotheses from her theoretical considerations, citing evidence in favor of these hypotheses either on grounds of existing

empirical evidence or by generating empirical (generally experimental) evidence herself[24]. Looking at social norms from a RC perspective already does produce valuable insights. Ultimately, then, it is its potential to offer such new insights against which the value of a RC account of social norms will have to be assessed.

References

Aumann, R. (1987), Correlated Equilibrium as an Expression of Bayesian Rationality, in: *Econometrica* 55, 1–18

Axelrod, R. (1984), *The Evolution of Cooperation*, New York

Binmore, K. (2005), *Natural Justice*, Oxford

Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Norms*, Cambridge

— (2010), Norms, Preferences, and Conditional Behavior, in: *Politics, Philosophy & Economics* 9, 297–313

— /R. Muldoon (2014), Social Norms, in: Zalta, E. N. et al. (eds.), *Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/archives/spr2014/entries/social-norms/ [March 01, 2015]

— /X. Erte (2009), Do the Right Thing: But Only if Others Do So, in: *Journal of Behavioral Decision Making* 22, 191–208

Bolton, G. E./A. Ockenfels (2000), ERC: A Theory of Equity, Reciprocity, and Competition, in: *The American Economic Review* 90, 166–193

Coleman, J. (1990), *Foundations of Social Theory*, Cambridge/MA

Elster, J. (1991), Rationality and Social Norms, in: *European Journal of Sociology* 32, 109–129

Fehr, E./K. M. Schmidt (1999), A Theory of Fairness, Competition and Cooperation, in: *The Quarterly Journal of Economics* 114, 817–868

Fudenberg, D./J. Tirole (1991), *Game Theory*, Cambridge/MA

Gintis, H. (2009), *The Bounds of Reason: Game Theory and the Unification of Behavioral Sciences*, Princeton

— (2010), Social Norms as Choreography, in: *Politics, Philosophy & Economics* 9, 251–264

Güth, W./R. Schmittberger/B. Schwarze (1982), An Experimental Analysis of Ultimatum Bargaining, in: *Journal of Economic Behavior & Organization* 3, 367–388

Hendrich, J./R. Boyd/S. Bowles/C. Camerer/E. Fehr/H. Gintis (2004) (eds.), *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, Oxford

Kliemt, H. (2009), *Philosophy and Economics I: Methods and Models*, München

Lahno, B. (2007), Rational Choice and Rule-Following Behavior, in: *Rationality and Society* 19, 425–450

— (2009), Norms as Reasons for Action, in: *Archiv für Rechts- und Sozialphilosophie* (ARSP) 95, 563–578

— (2010), Norms of Evaluation vs. Norms of Conduct, in: Baurmann, M./G. Brennan/R. Goodin (eds.), *Norms and Values*, Stuttgart, 95–112

24 See, e.g., Chapters 2 to 5 of her 2006 book and Bicchieri 2010.

Lewis, D. (1969), *Convention*, Cambridge/MA
Opp, K.-D. (1979), The Emergence and Effects of Social Norms, in: *Kyklos* 32, 775–801
— (1983), *Die Entstehung sozialer Normen*, Tübingen
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge
Paternotte, C./J. Grose (2013), Social Norms and Game Theory: Harmony or Discord?, in: *Brit. J. Phil. Sci.* 64, 551–587
Rabin, M. (1993), Incorporating Fairness into Game Theory and Economics, in: *American Economic Review* 83, 1281–1302
Schelling, T. (1960), *The Strategy of Conflict*, Oxford–London–New York
Sen, A. (1973), Behaviour and the Concept of Preference, in: *Economica* 40, 241–259, reprinted in: Sen, A. (1982), *Choce Welfare and Measurement*, Cambridge/MA, 54–73
Sugden, R. (1986), *The Economics of Rights, Co-operation and Welfare*, Oxford
Taylor, M. (1987), *The Possibility of Cooperation*, Cambridge
Ullmann-Margalit, E. (1977), *The Emergence of Norms*, Oxford