



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2009

The motive of commitment and its implications for Rational Choice Theory

Herfeld, Catherine

Abstract: This paper addresses the explanatory role of the concept of a motive for action in economics. The aim of the paper is to show the difficulty economists have to accommodate the motive of commitment into their explanatory and predictive framework, i.e. rational choice theory. One difficulty is that the economists' explanation becomes analytic when assuming preferences of commitment. Another difficulty is that it is highly doubtful whether commitment can be represented by current frameworks while (pre-)serving the 'folk-psychological' idea of what is commonly understood by the idea of a commitment. Both difficulties lead to the conclusion that, although motives do matter, conceptualizing the motive of commitment would cause trouble for rational choice theory

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-142304>

Journal Article

Published Version

Originally published at:

Herfeld, Catherine (2009). The motive of commitment and its implications for Rational Choice Theory. *Analyse Kritik*, 31(2):291-317.

Catherine S. Herfeld

The Motive of Commitment and Its Implications for Rational Choice Theory

Abstract: This paper addresses the explanatory role of the concept of a motive for action in economics. The aim of the paper is to show the difficulty economists have to accommodate the motive of commitment into their explanatory and predictive framework, i.e. rational choice theory. One difficulty is that the economists' explanation becomes analytic when assuming preferences of commitment. Another difficulty is that it is highly doubtful whether commitment can be represented by current frameworks while (pre-)serving the 'folk-psychological' idea of what is commonly understood by the idea of a commitment. Both difficulties lead to the conclusion that, although motives do matter, conceptualizing the motive of commitment would cause trouble for rational choice theory.

0. Introduction

This paper presents a critique of the behavioural foundations of rational choice theory (hereafter RCT). RCT is based on the idea that preferences and not real motives describe behaviour. We ask, however, whether people's motives could and should occupy a central role in explanations and predictions of individual and social behaviour. This question is discussed by focussing on the motive of commitment and its importance to explain norm-conformity. Thereby, "making choices based on one's commitments comes to be understood as a matter of making choices irrespective of any personal gain or loss" (Davis 2004, 4). The paper analyses the potential role of commitment as a concept in explaining and predicting behaviour. It discusses the difficulties and potential implications of conceptualizing such a motive for RCT.¹ The paper concludes that motives

¹ Note that we are not making any claims here about the origins of human motives, nor are we concerned with their cognitive or neurological character. When we talk about motives in this paper, we are referring to our commonsense understandings of motives as the activators of behaviour directed towards certain goals. A human motive is understood in various psychological theories as being attributed to specific reasons a person holds; thus, having a reason motivates an action. The nature of such reasons can be manifold. They can be traced back to desires such as being hungry which provides a reason to eat something. But motivation can also be attributed to less apparent reasons such as morality, altruism or personal commitments. In this case, a commitment provides the agent with a reason to act which is different in nature than more obvious sources such as bodily desires; the difference manifests itself in the behavioural pattern they bring about.

matter for explanation and prediction. However, conceptualizing a motive of commitment already presents a challenge for the rational choice theorist.

We will place this critique in the context of the work of Nobel-prize-winning economist Amartya Sen, especially focusing on his contributions to the discussions on the limitations and explanatory shortcomings of classical economic theory since the 1970s. In his famous paper *Rational Fools* (1977), Sen started questioning the two main requirements for rational agency: first, he criticized the basic consistency axioms required to be fulfilled by the agent's preference ranking. Second, he questioned the assumption of maximization that constitutes standard RCT, i.e. that a rational agent must always choose the most preferred option available to him.² Thereby, Sen initiated a debate about the status and role of RCT and its potential shortcomings to accommodate behaviour motivated in fundamentally different ways. His work on commitment raises questions concerning the extension or rejection of RCT as an explanatory and predictive tool in the social sciences.³ It also continues to influence discussions about questions such as whether RCT should be interpreted as an explanatory and/or a normative framework for accounting for human behaviour. As there has been more controversy than agreement upon these issues, Sen's work remains highly relevant in the social sciences and especially in economics down to the present day.

1. Background

In recent years, a lot of theoretical and empirical research has been conducted to explain different kinds of human behaviour by social scientific models based on the standard economic view of RCT. Many objections have been raised against the theory and its usefulness and status has been discussed endlessly. The picture drawn by the economist of the *Homo Economicus*, a utility-maximizer who is only motivated by the aim of satisfying his own preferences has been highly criticized as being one-sided and narrow in scope when it comes to explaining human behaviour in its entirety, especially with regard to cooperation and pro-social behaviour. While the theory predicted defection and free-riding, empirical results in experimental settings of the ultimatum-, the dictator-, the trust- and the prisoner's dilemma-game showed that individuals systematically cooperate at the expense of increasing their personal utility. So, one main objection, especially

² As will be discussed later in the paper, this requirement directly relates to the underlying concept of motivation in RCT; the satisfaction of one's own preferences motivates the agent. Doing this is assumed to be in the interest of the agent.

³ The most recent work discussing these issues is a book edited by Fabienne Peter and Hans Bernhard Schmid entitled *Rationality and Commitment* published in 2007 by Oxford University Press. It presents a collection of essays written by economists and philosophers for a workshop hosted at the University of St. Gallen in May 2004, addressing issues such as committed action, the concept of rationality in economics and the role of identity and (collective) intentions for economic theory in the context of Sen's work. An earlier reference to discussions on the motive of commitment and the difficulties of rational choice theory are papers of the 'Symposium on Sen's philosophy', published in the Journal *Economics and Philosophy* 17(1). The content of both publications partly overlap.

raised by behavioural decision theorists, was that models based on traditional RCT could not accommodate this data (Camerer/Fehr 2002; Fehr/Fischbacher 2004).

Facing this objection, a variety of reductionist arguments have been given (Jolls/Sunstein/Thaler 1998). Economists have extended the key concepts of constraints, preferences, and utility in their ‘as-if-approach’. They broaden their understandings of the assumptions of maximization and self-interested preferences. They allowed for other kinds of preferences to accommodate behaviour originating from a variety of motives.⁴

In the same line, there has been a recent trend in the social sciences to explain behavior by referring to the existence of a social norm, which guides people’s behaviour and motivates them to cooperate in certain (social) settings.⁵ This seems plausible; empirical data is indicative of the fact that people conform to social norms to interact with others. Fairness, reciprocity, and promise-keeping enable or enhance cooperation amongst the members of a given group or society. Norms are supposed to be solutions to market failures and to collective action problems (e.g. the free-rider-problem) (Arrow 1971; Ostrom 2000; Coleman 1990; Ellickson 1994). Consequently, despite the problem that social norms are generally a) hard to measure and b) only vaguely understood in terms of their actual influence on people’s behaviour, they have been increasingly acknowledged by social scientists and economists alike. In the case of economists, manifold attempts have been made to integrate norm-conformity into the rational-choice framework and its concepts (e.g. Akerlof 1982; Bicchieri 2006; Conlin et al. 2003; Coleman, 1990; Fehr/Fischbacher 2004; Henrich 2000; Basu/Weibull 2003; Kahneman et al. 1986; Lindbeck et al. 1999; Elster 1989). Hence, social norms play an increasingly significant role in the explanation and prediction of cooperation in all the social sciences. The wide acknowledgement of a social norm as a concept and the extensive experimental literature on norm-conformity makes norm-conformity a good example for evaluating the behavioural foundations of RCT, for analysing the explanatory role of motives and for discussing the difficulties and implications of something like commitment as a potential motivating force for behaviour such as norm-conformity.

The works of Bicchieri (2006) and of many other scholars in this field have to be acknowledged as highly influential and insightful. However, one can ask whether models based on RCT can provide useful and extensive explanations for all kinds of individual and pro-social behaviour. This question arises, because, independently of the characteristics of whatever motive assumed, all actions ex-

⁴ Besides self-interest in which case a positively monotonic utility for one’s own material payoffs and indifference about others’ material payoffs is assumed, they acknowledged other-regarding or social preferences that are characterized by utility that is not constant with respect to variations in one’s own or others’ material payoffs (Fehr/Fischbacher 2002). For example, altruistic preferences are characterized by utility that is monotonically increasing in others’ material payoffs and one’s own payoffs; and emotions can be modelled as positive or negative internal sanctions, which affect the utility function of the agent in the form of psychic costs or benefits (Elster 1998).

⁵ We will call this behaviour norm-guided behaviour here. It should not be confused with cooperative behaviour in general, although in many cases by following social norms, people can be said to cooperate with each other.

plained by RCT are ultimately reduced to instrumental agency. In the end, the agent is always assumed to seek the satisfaction of his own preferences.⁶ And it is its specific character which provides us with another reason why ‘norm-conformity’ serves as a good example to evaluate RCT as an explanatory and predictive tool: we think that norm-conformity can have an instrumental, but also a normative dimension. This is because social norms not only provide solutions to problems of cooperation. Social norms also make prescriptions about acceptable and unacceptable behaviour within a group or society. We believe that, depending on the motives a person has (e.g. cognitive motives such as commitment or duty and non-cognitive motives such as emotions),⁷ actual actions are instrumental and/or normative in character. And we think that, in contrast to the ‘instrumental dimension’ which is well captured by RCT, it is the ‘normative dimension’ that has not yet been successfully accommodated within RCT.

2. Line of Argument

Having this background in mind, firstly, we argue that there is a dimension to agency that is heavily underestimated by RCT, which is what we will call the ‘normative dimension’ of agency. By this we mean that human beings are able to consider and reflect on what they ought to do, independently of whether this ought springs from an internal judgment about right or wrong (behaviour) or whether it has its origin in the expectations and judgments on the part of society. This normative dimension requires the ability to consciously adopt and abide by certain norms and thereby to distance oneself from one’s own bundle of preferences. This requires as person to have the capacity to undertake judgments about right or wrong, independent from immediate desires or inclinations. It requires to base or adjust one’s choices and actions on these judgments and to have the capacity of self-determination/formation, self-control and ultimately of forming one’s (personal and social) identity, and perform actions in line with these judgments. The economic perspective denies and even contradicts

⁶ In this paper, the concept of welfare directly relates to the satisfaction of personal preferences and utility. This view of welfare is also called “preference satisfaction view” (Hausman/McPherson 2006, 120). It takes utility-maximization as leading to welfare-improvement. However, the two concepts of individual utility and welfare should generally not be confused with one another or considered as synonymous. In the case of social or altruistic preferences, an individual can maximize his individual utility by satisfying his altruistic preference which improves, however, the welfare of another person. In the case of social preferences, the improvement of social welfare is the result of satisfying individual preferences. Consequently, maximizing personal utility by satisfying individual preferences does *not* necessarily imply the improvement of personal welfare.

⁷ The classification of motives as well as their typologisation depends on the kind of psychological theory one want to adopt. For reasons of simplification, we will make the rather amateurish claim that there are three types of motivational sources under which many motives can be subsumed; that is self-interest, emotions, and what the person voluntarily chooses to on his or her better judgment against his or her immediate inclinations (e.g. doing s.th. out of duty, commitment). This goes in a similar direction as Mansbridge 1998. She distinguishes between narrow and wide self-interest, duty and love as the three main types of driving forces behind human behaviour which bring about distinct patterns of behavior.

this dimension. However, this normative dimension of agency can be essential in motivating individual and pro-social action. We argue that being able to explain why people keep conforming to social norms over long periods of time, why they vote or stick to religious practices even if their potential benefit remains completely uncertain requires taking seriously people's motives beyond preference-satisfaction and utility-improvement. We suggest that the motive of commitment, famously characterized by Amartya Sen (1977) and later developed in a more sophisticated way by other scholars (Peter/Schmid 2007), could help us capturing the normative dimension of agency and the type of motivation behind such behavior, here norm-conformity.

Secondly, the underlying rationale of making decisions based on one's judgments is the idea of maximization and the satisfaction of individual preferences in RCT. However, the rationale for choices which have a normative dimension is shaped by what the agent values rather than by what he prefers. We argue that therefore a motive of commitment cannot be conceptualized within the RC-framework and that the RC-perspective rather contradicts this idea. Economists have suggested assuming an unconditional preference for complying with commitments or modelling commitment as constraint on the utility-function. However, according to the common 'axiomatic interpretation' of RCT and the theory of revealed preferences accompanied with it, this entails indeed a mathematical formulation, but does not provide us with empirically testable hypotheses and causal relationships.⁸ Assuming a preference as explanans for the behavior we want to explain, i.e. the explanandum, this leads to the analyticity of our statements rather than to proposing testable hypotheses. This again weakens the explanatory and predictive power of RCT and invites for questioning its status as a useful tool in scientific research.⁹ Modelling commitment as constraint on the utility-function can be rejected on the grounds that we want our concepts to preserve, at least approximately, our 'folk-psychological' idea of what is commonly understood as a commitment and the nature of an action commitments bring about. Hence, the motive of commitment cannot be accommodated by RCT.

However, it is argued in this paper that it is important to distinguish between varieties of motive-types when we take RCT as an explanatory tool; they are distinct in the sense that they bring about different behavioural patterns. Hence, in order to get adequate explanations of individual and social behaviour and make reliable predictions, the distinct nature of motives has to be acknowledged. By integrating other-regarding preferences and emotions into RCT a first

⁸ Heiner makes the point that the ultimate extension of the RC-view is not that agents maximize, i.e. select the most preferred options/actions but rather that "any observed behavior is consistent with the maximization of some function" (1983, 561). The latter is incapable of theoretical or empirical disproof (Boland 1981); it is rather a mathematical fact and thereby difficult to reject on solid grounds.

⁹ We adopt Sen's view of a 'useful explanation' in this context. He argues that scientists do neither aim at an "assumption structure [within a theory] which is fundamentally at odds with the real world, nor at one where simplicity takes the form of naivety" (Sen 1987). This differs from an instrumentalist picture where the behavioural assumptions can be unrealistic insofar as the theory provides good predictions.

important step has been already made. The concept of commitment would be a further motivational force which would help to account for the normative dimension of agency in individual and social life. If RCT would access something like a typology of the most important motives, we will attain a better understanding of the role motivation plays in the explanation of human behaviour.

To develop our argument we shall first recapitulate what RCT is and how RCT is used to explain why people (think they ought to) follow a social norm. We will tackle the question of what the nature of normativity involved in norm-conformity is and how it relates to motivation. We conclude that such behaviour can have two dimensions, an instrumental and a normative one, whereby the latter has difficulties to be accommodated in models based on RCT. By providing a characterization of the motive of commitment, we will show why a conceptualization of such a motive could play a role in accounting for the normative dimension of agency in cases we want to account for this behavior. We will then show the difficulties of including such a concept of commitment into models based on RCT. Finally, we examine the conceptual and methodological implications of acknowledging the concept of commitment.

3. Rational Choice Theory as an Explanatory Framework

To address the question of whether people's motives could and should occupy a central role in RC-explanations and predictions, we have to understand how behavior in general and norm-conformity in particular is explained by models based on RCT and what the economist considers a social norm to be.

3.1 Introducing Rational Choice Theory

As its name suggests, the basic idea of RCT is that human beings choose rationally. Specifically, the framework is based on the behavioural assumption of instrumental rationality, i.e. the agent's sole motive for action is assumed to lie in the satisfaction of personal preferences at the expense of other possible motivations (Sen 1985). Individual preferences as well as individuals' beliefs about specific events are given. The intuitive principle of a rational choice is that individual agents choose those actions which they expect will have the best consequences for them, given their beliefs and desires. Putting this idea in a more formal principle of rational choice, it states that rational agents choose the action with maximum subjective expected utility.¹⁰ The idea behind both principles

¹⁰ The term 'utility' is often seen as representing the person's welfare or well-being (Sen 2002, 27). This is misleading in at least two ways: first, it sounds as if the agent is rational in the moment he aims at maximizing his personal welfare in the first place. It suggests a selfish agent, who is only focused on what lies in his immediate personal interests (his self-interest), or personal welfare, and what is in his interest only relates to himself. Interest in other people, as the word 'self' already implies, is excluded on this interpretation. Second, welfare often gets confused in the literature with the definition of 'utility' as the *maximand* of a person's choice behavior that is inspired by the theory of revealed preferences (Sen 2002, 27). But in ordering his preferences the agent is assumed to attach utility to the options, not 'welfare-units'.

is that choosing rationally entails choosing the option, or a strategy in game theory, which will best satisfy the individual's preferred ends, given a finite set of alternatives (or, in game theory, a given choice situation). What best satisfies the agent's preference is the option that has the greatest utility to the agent. Consequently, a performed behavior is a form of optimizing behavior, taken to satisfy the agent's personal preferences and thereby maximize individual utility (Elster 1989).¹¹

3.2 Intentional Explanation

When rational choice theorists explain action, they focus on observable behaviour. This is because, according to the standard economic reasoning, we cannot observe people's motivation for performing a certain action. What goes on in people's heads is taken to be inaccessible and the *true* motivational force remains obscure. The economist's explanation is rather based upon what he generally assumes to be characteristic for a motive of action. Motivation is grounded on the concept of intentional action used in folk psychology. An intentional action is a "piece of behaviour" which derives from the intention of the individual exhibiting it (Elster 1994, 311). If an agent acts instrumentally rational, that agent's intention is what is assumed to be motivating that very action. The intention of the agent is represented by a reason an agent holds for the performance of an action to achieve a certain goal. This is what distinguishes action from mere behaviour in RCT, namely that it is always done with a purpose (e.g. Anscombe 1957; Rosenberg 1995). Reasons are modelled as being provided by mental states, i.e. the desire to do something and the belief that a certain action serves as a means to satisfy this desire (Elster 1994; Hausman 2005; Rosenberg 1995). For the economist, the motive for an action is thus taken to be identified by providing a reason, i.e. a desire and a belief which the agent had to perform an action. The observed behaviour is the resulting outcome or effect of a causal process between two events, the reason and the corresponding action. So the motivational structure of an action is assumed as causal in nature when providing an explanation

¹¹ This is a very delicate point that often gets confused and thereby leads to misunderstandings and false interpretations of the concept of rational choice. The agent does not aim at maximizing utility in the first place. He is motivated to act because he wants to satisfy his preferences. And *because* the agent prefers one option x_1 to another x_2 , the utility of x_1 is higher than the utility of x_2 , not vice versa. And this is why, in choosing x_1 , the agent also gets the highest utility, namely U_1 . So, utility-maximization is not the reason for his action but a consequence of his individual preference-ordering. As preferences cannot be observed but we assume that the agent takes the option he most prefers, we infer some information about the nature of agent's preferences from the observed choice. This view provides the basis for the well-known view of revealed preferences (Samuelson 1938; Sen 1977).

(Davidson 1963).¹² The observed behaviour is then explained by attributing a reason as its cause.¹³

In the vocabulary of standard RCT, the desires of the person are translated into a preference-set. The beliefs a person holds are taken to be the set of information available to the agent. Formally, preferences are represented by a utility function and beliefs are represented in terms of a probability distribution; this is in case that it is assumed that the agent does not have full information about the consequences of all options available to him. To evaluate or choose the right means to achieve his ends, the agent is assumed to have a rank ordering of the options available to him according to his preferences. They are measured in terms of 'utility' and the ordering is represented by the (expected) utility attached by the agent to the different (material or immaterial) outcomes. Aiming at the satisfaction of his preferences, he chooses the option which is assumed to provide him with the highest (expected) utility and acts accordingly, given his preferences and information available. The results of the actions are the objects of desires or, better say, objectives that satisfy the agents' preferences in the best possible way. Accordingly, the satisfaction of preferences is assumed to always be the person's main end and is therefore said to be in his/her self-interest (Hargreaves-Heap et al. 1992). The limited empirical basis for unobservable preferences or desires does not present a problem in this context. Instead, it is simply claimed that the observed action reveals the preference. And preferences reveal the motivations involved. According to this understanding, different preferences (re-)present different motives; and as preferences are given in economic theory, motives are not something the economist is concerned about.

Note that there is an abundance of literature on the topic that has spawned a variety of very different interpretations and uses of RCT.¹⁴ Therefore, it is difficult to pin down exactly what has been and what can be legitimately

¹² Note that it remains opaque and also irrelevant for the explanation in this picture what the ultimate reason of a person is. This relates to an often raised point against the view that motives should matter. It is often claimed that finding out about people's motivation is too difficult; that even the person himself might not be able to identify what really motivated him. Or that there might be a chain of reasons which jointly caused the action in question (Rosenberg 1995).

¹³ In the RC-literature, there is no difference made between choosing a course of action or choosing an option and the actual performance of an action. It is assumed that the choice of the agent based on his preferences necessarily leads to the performance of an action or that the choice is the actual action. This is not obvious but cannot be discussed here. However, what matters later in the paper is that a choice which I would make based on my preferences does not lead to the appropriate behaviour, because what my preferences dictate can conflict with the course of action that my commitments dictate. And because I want to act on my commitments I perform an action in line with my commitment rather than performing an action that would satisfy my preferences. This is one reason why RCT has problems to accommodate committed action. Commitment separates choice from utility-maximization and preference-satisfaction. This is important and we will come back to this aspect later on in the paper.

¹⁴ Bunge (1996) makes some attempts to disentangle the various versions of RCT and the various interpretations of its concepts. For example he finds more than 10 interpretations of the concept of rationality made within the context of rational choice theory. He claims that rather than being one theoretical framework, RCT rather consists of a family of theoretical models and concepts. He also highlights the differences between contemporary rational choice theories and their precursors. This should give a very rough idea of how diverse the different

subsumed under the label of RCT in the social sciences; and to what extent social scientists' objections to RCT as it was originally conceived in economics applies today. Even in mainstream economics, the term 'rational choice' has been used in many different ways.¹⁵ However, at least since the 1950s, 'rational choice' has been associated with a set of core axioms that describe it as being a consistent choice. A consistent choice requires a set of *a priori* conditions of internal consistency to be fulfilled by the agent's observable choices in different situations where the agent chooses between different menus, i.e. different sets of options available to him. The exact consistency-requirements can vary, but the standard ones relate to the idea that a set of actual choices can be explained as resulting from maximization according to some binary relation which represents the choice function or utility function of the agent (Sen 1987). This relation is required to be transitive, reflexive and complete. However, consistency alone does not imply any kind of normativity of the agent's intentions. It does not entail that either the performed action has valuable content for the agent besides fulfilling its instrumental role or that the end is perceived as *right* in a more profound sense (Raz 2004)—*right* in the sense that the agent judges it so. The normativity involved is of an instrumental nature. If an agent is rational, he has to choose according to what his preferences dictate. There is no space for the agent considering the action itself as worth pursuing, independently of the fact that it satisfies his personal preferences. What the agent chooses as the right action is assumed to result exclusively from a judgment based on the rationale of preference-satisfaction and maximization. This can be taken as the standard framework of RCT.

3.3 Rational Choice Theory and the Concept of a Social Norm

There is no unified definition of a social norm used by social scientists (Hechter/Opp 2001; Stout 2001; Okruch 1999). However, there are some key features or properties generally considered to be common to social norms. First, social norms are characterized as informal rules of behaviour "that people follow for some reason other than the fear of legal sanctions" (Stout 2001, 6). This means

schools of thought of the different social sciences are when it comes to break RCT down into one category.

¹⁵ The picture we have just outlined here refers to Amartya Sen's view, who claims that RCT can be identified with a "specialized school of thought" (2002, 11). This school represents the standard view of modern economic theory that social scientists have in mind when they refer to the concept of rational choice in the context of RCT, game theory, and decision theory. Depending on the interpretation of concepts such as desires/preferences, constraints, utility, rationality and decision/choice, which additional assumptions are accepted and for which purpose they are being used, the underlying notion of rational choice can change dramatically in both meaning and character. However, rationality as internal consistency that is secured by the fulfilment of these very abstract axioms, allows for a high flexibility in interpreting its concepts. But very frequently, economic models still assume that individuals are self-regarding. Take for example consumer theory, which assumes that individuals do not care what others are consuming but only making choices on the basis of what the individual himself prefers. Besides that, RCT is also not committed to any particular explanation of the origin of an agent's preferences and especially to a position on whether these preferences have social or biological roots.

that they are not externally enforced by formal rules, such as legal norms, nor through coercion. Instead, they are usually enforced by informal sanctions executed by third parties, e.g. such as the exclusion from a social group by its members (Mantzavinos 2001). Or the norm is enforced by mechanisms internal to the agent, i.e. negative or positive sanctions. This is in case he feels a strong obligation to conform to the norm, because he internalized it through the process of socialization, but does not prefer to do that, in which case the transgression will cause negative feelings such as guilt.

But social norms do not have to be enforced at all. Social norms can become part of our system of values (Bicchieri 2006). In this case, the agent wants to act in line with the norm as he values what the social norm embodies. For example, in the case of fairness, the agent consciously may want to conform to the norm of fairness, i.e. to share equally because he values fairness and because this is what the norm stands for. Another feature is that social norms are ‘social’ in the sense that they are public, shared and accepted by the members of a group or society (Hechter/Opp 2001; Fehr/Fischbacher 2004). Although the specific set of norms can differ cross-culturally, social norms are assumed to exist in every society, every subgroup of society and any kind of organisational structure (Kirchgässner 2006).

In the RC-literature, social norms are often functionally defined by the interest they serve (Hardin 2008). The motivation for norm-conformity is modelled as deriving from people’s preferences. The agent has a certain end he wants to achieve and he considers conformity to a specific social norm as the best means to achieve this desired end within his set of options. On this view, norm-conformity is interpreted as a form of optimizing behaviour; it is assumed to be behaviour performed to satisfy one’s preferences and thereby maximize individual utility (Elster 1989). Conformity then involves choice from the agent’s part; the utility of the agent is *directly* affected and it is contingent upon the agent’s preferred goals. Hence, if the preferences of the agent change or if the agent has a more efficient means to satisfy his preferences, he may substitute norm-conformity by a different course of action. Norms of etiquette may increase an agent’s chances for getting a job in a prestigious firm, the achievement of which would satisfy the agent’s preference for that job. The goal of getting a job motivates the agent, to follow the norms of etiquette. Avoiding (informal) costs of sanctions expected by the agent in the case of transgression can also satisfy the agents preferences; in this case, personal utility is *indirectly* affected through (positive or negative) emotions or external but informal (positive or negative) sanctioning in case of conformity or deviation respectively.

On this view, social norms are in a way to be (value-)neutral (Hardin 2008). The normativity involved in such norm-conformity, i.e. the reason why people think they ought to follow a social norm, is purely instrumental in character. An instrumentally rational agent thinks he ‘ought’ to follow a social norm because it is in his interest to do so. His internal motivating source for conformity rests on a prudential or—as we would call it—‘instrumental reason’ to obey the norm

(Opp 1999; Coleman 1990).¹⁶ In the following we will refer to ‘instrumental normativity’ or the ‘instrumental dimension of agency’ in this context.¹⁷

In contrast to this view, imagine that what is pro- or prescribed by social norms can conflict with what one’s own preferences dictate in a given situation. Conformity can be motivated by a goal that does not satisfy one’s preferences and utility-maximization. This is why often some additional motives are required to be in place to make social norms efficacious. Norm-conformity can have many different motivational sources. Emotions or commitment can play a crucial role in norm-conformity (Elster 1989; Mansbridge 1998; Zimbardo et al. 1996). We suggest in the following section that the normativity involved in the norm-conformity can fundamentally differ from instrumental normativity described above. Whether or not it differs depends on the motivation a person has for conforming to social norms.

4. Two Dimensions of Behavior and the Explanatory Role of Motives

In this paper, we are concerned with the behavioural foundations of RCT. Originating from economics, RCT is a framework that is extensively used in all the social sciences to account for a variety of behaviour in different social settings. However, we have indicated that RCT cannot be used to explain all kinds of behaviour. To justify this claim, we have suggested to analyze norm-conformity as one example of behaviour where RC-explanations are sometimes questionable. To do that, we have introduced social norms and how they are characterized in the literature and we have outlined how a RC-explanation of norm-conformity would look like. We also have seen that RCT makes methodological assumptions of universal and objective scientific laws about human behaviour based on specific consistency-requirements; it draws a picture which favours individualism and takes rationality to be instrumental in nature. The two main assumptions about rationality underlying RCT are first, that behavior is outcome-orientated (in terms of maximization) and consequentialist in nature. Second, behavior is assumed to be instrumental in character which basically means that an action is

¹⁶ Bicchieri defines a prudential reason for action as “if you have a goal x and the best available means to attain x is a course of action y , then you ought to adopt y ” (2006, 14). This is similar of having what we called ‘instrumental reason’ as an instrumentally rational agent.

¹⁷ Note that normativity is not an issue which can be clearly contrasted with instrumental. Actions which have the instrumental dimension we are talking about here can also be taken as normative in character. The normativity involved in an instrumental action derives from the rational choice theorist’s understanding of rationality as being instrumental in character. On this view, the normativity involved in an instrumental action, i.e. why a rational agent ought to take one rather than another action is independent of any talk about values. It is rather determined by the conditions of how instrumental rationality is defined in RCT. However, this is different from the normativity involved in an action which has this normative dimension stated above. What we mean by normative here (in contrast to instrumental) is that when an action has to be taken because there are prescriptions/proscriptions of a norm external or internal to the agent but which provide the agent with independent reasons which he accepts or endorses or which are in line of what the agent himself considers as right/wrong, independent of he is concerned about in this situation.

performed just because it is a means to achieve another goal. Both assumptions also hold for explaining norm-conformity on this view. However, we have also indicated, that social norms pro- or prescribe actions which can conflict with what individual preferences dictate in a given situation.

RCT seems to contradict this normative dimension of agency. Three questions require an answer: when does behaviour have this normative dimension and why is such a flexible approach not be able to account for it? And if so, what could we do about this ‘explanatory gap’? In the following, let us look at each of these questions in turn. But first let us understand what we mean by the normative dimension of behaviour by exemplarily looking at norm-conformity.

4.1 The Normative Dimension of Social Norms

We want to say that the underlying problem of explaining norm-conformity with RCT is that norm-conformity seems to involve an instrumental *and* a normative dimension. What does that mean? A quick look over the literature allows us to pick out overlapping notions such as ‘expectations’, ‘obligatory’, ‘value-judgments’, ‘right’ or ‘wrong’ and ‘authoritative’ when it comes to defining social norms (Anderson 2000; Bicchieri 2006; Coleman 1987; Elster 1989; Hechter/Opp 2001). These normative notions suggest a type of normativity attached to norm-conformity which imposes requirements on supposedly rational behaviour that are different from the minimal consistency-requirement postulated by RC-models. From this point of view, norm-conformity is neither regarded as being a mere regularity of behaviour.¹⁸ Nor is it considered to be completely contingent, either upon people’s preferences and expected utilities, or upon the usefulness of the norm in a given situation.¹⁹ Instead, social norms are often taken as more or less stable “*prescriptions* that establish how one *ought* to behave” within a community (Woodward 2009, 33, my italics). They are defined as “standards of behaviour that are based on widely shared beliefs how individual group members *ought* to behave in a given situation” within the community (Fehr/Fischbacher 2004, 185, my italics). They are said to “reinforce certain patterns of behaviour [...] by representing these patterns as desirable or *obligatory*” within a community (Pettit 1990, 725, my italics). In this sense, a social norm is a rule prescribing or proscribing behaviour; a rule that can be represented by a statement of the form of a conditional, hypothetical imperative ‘in situation X, you *ought* to do Y’. In this case, they are conditional upon the circumstances and not, as instrumental rationality assumes, exclusively on outcomes. Or they have the form of an unconditional, categorical imperative ‘you should (not) do

¹⁸ Bicchieri (2006) contrasts social norms with descriptive norms such as fashions and fads, whereas the latter are mere regularities which can be observed. According to Bicchieri and in line with my argument, what classifies collective behaviour as a descriptive or a social norm is a) the motives of the people involved and b) the expectations of the people within the community.

¹⁹ There is a lot of empirical evidence that conformity of an individual varies from situation to situation. But it seems undeniable that which kind of social norm is being used in a specific context is stable over time and counts for every individual equally. Table manners are supposed to be conformed to for example when people have dinner together. This is not to say that each individual does follow the appropriate norm in this situation.

X' in which case an individual conforms to a social norm for the sake of the norm independent of the consequences (Elster 1989, 98). The non-consequentialist character, the conformity for the sake of the norm itself or for what the norm stands, and the conformity independently of individual preferences are aspects that are meant to be captured by the 'normative dimension of agency'.

We claim that whether or not behaviour has a normative dimension in the abovementioned sense depends on people's motivation of their behavior. People's motives for conformity can be manifold (e.g. Bicchieri 2006; Camerer/Fehr 2002). People conform to social norms because they fear what others would say in case of transgression. People want to live up to other people's expectations, or they do not even know that they conform to norms; they often just internalized them and follow them unconsciously in form of a habit or custom (Bicchieri 2006). These types of behaviour are instrumental in kind; here, conformity has a goal that is different from conformity itself, but which is preferred by an individual. Achieving his goal presents the agent with a reason to conform to the norm.

Sometimes, however, individuals consider the conformity to social norms as normative for them, because of what the social norm embodies; they are motivated to conform to the norm because they think that the norm supports bringing about the appropriate kind of behaviour. Especially social norms of cooperation such as promise-keeping, fairness, truth-telling and reciprocity are operative by supporting value-laden patterns; they align people's behaviours with their values. And we often, we follow the norm to live up to what we take as valuable. If fairness is considered as valuable by an individual, it takes the norm of fair sharing as the right thing to do as the norm reflects what he values. Conformity does not serve as the best means to attain the optimal outcome for the individual in a given situation. It is rather that the individual values fairness and this provides him with a reason to follow the norm, independent of his personal goals and potential instrumental reasons the agent can also have. He follows the norm even if conformity causes him to be personally worse off.

This is so because social norms often set standards of behaviour which are derived from an accepted set of values.²⁰ They are standards that provide a basis for individual judgments of options in a choice situation.²¹ Social norms often represent or embody what people value in a society. And people want to align their values with their behaviour and social norms can serve this function in case they embody a value the agent himself holds (Wallace et al. 2004). This is one reason why they often use the social norm as an underlying rationale to

²⁰ These values can be for example social, cultural, moral etc..

²¹ These standards, i.e. the social norms themselves, have to be at least 'passively' approved of, if not actively ratified, enforced and followed as well as implemented by the individual. 'Passively approved of' in this context means that the existence of a norm is accepted even if people avoid situations where the conformity to a social norm would have been expected, like the example of the hunter-gatherer community of the Ik's shows (Turnbull 1972). Thus, even if people do not follow the social norm, its existence is common knowledge and thus not denied and can therefore be assumed to be passively accepted. The important point here is that such a collective acceptance requires a shared understanding of what is regarded as socially appropriate or inappropriate among the individuals within a community (Krupka/Weber 2008). This common understanding is achieved on the basis of a generally accepted set of shared values.

make judgments and rank their options and for making a choice of a course of action.²² What they value thereby shapes their choices; they are not shaped by what they prefer in these cases.²³ Because the underlying rationale for making judgments of options represents the individual's motives, by using social norms as evaluative standard, values can become part of what motivates the individual to choose. Like the agent's option-ranking according to their preferences leads to an action they have to take if they are instrumentally rational, values can also make the performance of a kind of behavior normative, i.e. necessary for an individual.²⁴

Motives shaped by preferences and motives shaped by values are quite distinct in the sense that two different rationales are consulted by the agent for making his judgment on which he ranks his alternatives and bases his choices.²⁵ In the latter case, individuals transform their value-orientation into their individual and collective actions through choice. Here, the agent 'reveals' his values by his conformity.

So there seem to be two dimensions of behavior which are distinct in the sense that the behaviour is motivated in different ways; necessitating choice and thereby entailing different reasons for action. A course of action is taken either because the agent bases his choice on what his preferences are dictating him. Or it is taken because an action embodies what he values. In the case of social norms, people's conformity to them can be one way of revealing what the individual considers as being the right thing to do. This can conflict with behaviour dictated by his preferences.²⁶ This is important because it shows the link between people's judgments over options and their motivation that leads to make choices and the performance of behavior. As the performance of behaviour is conditioned upon the motives, and motives are shaped by preferences and/or values, this is why motives determine the dimension of behaviour. And if we want to capture both dimensions, motives matter in explanation.²⁷

²² Another reason why values influence choices is that people often take values to be constitutive for the self-conception of their group, of themselves, and of their self-understanding as autonomous agents (Elster 1989). Beyond mere approval, the individual can even identify with certain values that are embodied in social norms. In this case, one could say that people feel committed to these values. Conforming to a norm makes it possible to reveal the agent's commitments and identity.

²³ It is fundamental for the argument here that preferences are assumed to be distinct from values. This is, however, not a distinction drawn in mainstream economics. For an interesting discussion of values in economics see e.g. Klamer 2003.

²⁴ This assertion should not be understood as a claim for a causal relation in this context. Necessity in this sense is to be understood as a requirement of rationality or rational agency.

²⁵ Only to remind the reader: the underlying rationale for making judgment and ranking options in standard RCT is assumed to be maximization and preference-satisfaction.

²⁶ It can be possible that the behavior which is dictated by one's preference happens to be the same course of action that represents the agent's values. For example the norm of fair sharing can improve my utility in a given situation. At the same time it is the case that I am committed to and identify myself with fairness as a value which makes it necessary for me to share fairly in every situation independent of the fact that in a particular situation it also enhances my personal welfare.

²⁷ We want to stress that the type of normativity, i.e. the necessity to take one course of action rather than another cannot be derived from the consistency-assumptions in this case

So, why does RCT not accommodate this kind of behaviour? Taking up on this idea of a social norm being a standard of evaluation for approving behaviour, the understanding of why people conform to social norms differs significantly in these cases from the notions of agency in standard economics. It would be useful to distinguish between the two dimensions by looking at motives. There is the ‘instrumental dimension’ deriving from the intuitively plausible consistency requirements; this type is well captured by RCT. The second is a ‘normative dimension’, which cannot be captured by RCT because the agent does not reveal his preferences by making consistent choice in the assumed sense. The agent rather reveals what he considers as right or wrong based on values he holds as a member or a group or society. Depending on people’s motivation, conformity is normative for an agent in either one or both of the two senses at the same time. It depends on the rationale they use for making choice judgments. If individuals evaluate the option ‘norm-conformity’ in terms of appropriateness or inappropriateness, they take their values or the social norm as a standard for evaluation for making their judgment. Again, in this case, to follow a social norm is normative for a person, i.e. the person recognizes and accepts the normative claim that that social norm makes on him/her, because the norm embodies the values the individual or group is committed to. And the “acceptance of a certain normative consideration can mould the way [one] thinks” (Millar unpublished, 1) and acts. What motivates people is indirectly the recognition of the claim that the norm makes on them and directly their commitment to the values embodied in that norm. Against the economists’ view, personal welfare-improvements and preference-satisfaction are not the reason for the choice of the agent; they are not what the agent aims at in the first place and thus are not what makes norm-conformity normative for them (Sen 1977; Hausman 2005). Norm-conformity is not instrumental in character, but normative in a different sense than demanded by the consistency-requirements for instrumental rationality in RCT.

We suggest that the difficulty of RCT to accommodate an action chosen by judging it as right or wrong, rather than judging it as the most efficient means to satisfy one’s preferences, can be resolved by introducing Amartya Sen’s concept of commitment. As commitment has not been extensively covered in the RC-literature, we try to give a rough characterization based on commonsense understandings.²⁸ By looking at Schelling’s and Sen’s accounts, we will indicate what a conceptualization and formalization of a motive of commitment would have to capture. And we will argue that such a concept could play a crucial role to accommodate the normative dimension of agency.

because the action does not necessarily reveal consistent behavior in the sense assumed by RCT.

²⁸ We have to admit that an underlying premise of the claim is the empirical statement that commitment as it will be characterized in this essay does exist. However, to ground this claim in empirical evidence is a difficult task. There is a lot of empirical research still to be done on filtering out whether there is such a motivating force as commitment and, if yes, what its sources are and to disentangle this from other motives. So far, we will just assume that there is a motive of commitment. For references to empirical evidence see Peter/Schmid 2007; Sen 2005 and 1987; Kerr et al. 1997 and Kerr/Kaufmann-Gilliland 1994.

5. The Concept of Commitment

We have argued that, depending on why people think they ought to perform a certain action and related to the reasons they hold for acting in a certain way, an action has an instrumental and/or a normative dimension. Acknowledging the importance of these two dimensions matters, when we aim at explanations and predictions of individual and social behaviour. Commitment as a motive reflects the main aspects of this normative dimension of agency and hence, could be very helpful in accounting for the normative dimension of agency.

5.1 What Commitments Are

When we think about commitment, we automatically think about something that constraints our choices or that brings with it a certain limitation of the possibility to pursue what we want. We do not necessarily think about something that we desire to do (Schmid 2007). They do not express what the agent *wants*. Rather, if a commitment presents a reason for action this is because commitments, as Searle puts it, create “desire-independent reasons for action” (Searle 2001). We have the intuition that commitments might even exist only for the purpose of keeping us in line with what we think would be better for us, were we to distance ourselves from our immediate desires to evaluate our options; they rather reflect what we generally value, instead of what we have a taste for in a certain moment.²⁹ This clearly reminds us of what we have said about the normative dimension of agency.

Commitments are often voluntary. Even if we feel them as constraining our set of options and thereby restraining our choice possibilities, these limitations are, at least in many cases, self-made. People voluntarily marry; they perceive themselves as members of a political party or a religion. They are committed to the company or organization they work for; they pay taxes in the country they live in; they stick to the realization of ideals such as ‘sustainable development’ or communism; they go to strikes and demonstrations to realize a common goal which often, if at all, only indirectly benefits them. People go out and vote, knowing that their vote will not have a direct effect on the outcome.³⁰

Generally speaking, commitments have an intentional object. They are always directed towards something, an *object of commitment* as we will call it and require a judgment subsequently to choice; a judgment in terms of what the

²⁹ An example would be the individual’s upholding or pursuit of the group interest at the expense of what the individual itself prefers. Collective action for the provision of the public good is achieved, although the individuals have a strong incentive to free-ride, given their preferences.

³⁰ The so called ‘value voters’ in the US is a large and important group. Their voting behaviour derives from their string commitment to the religious values they hold. Their choice is based on judgments about what they consider as right or wrong, which manifests itself in the agenda of one party rather than another. It is not manifested in what satisfied their preferences in the moment of the election. Note that this is different from the expressive voting account in which people express their support for one party rather than another. It is also different from the mainstream instrumental account, on which voting is taken as a revelation of preferences over possible electoral outcomes.

agent values, not in terms of what best satisfies his preferences and maximizes utility. Many things can serve as an object of commitment. As indicated above, it could be an ideal, a specific goal, or a specific moral or non-moral principle. What matters, is that our commitments are always about an object and that, depending on the nature of the object of commitment, our commitments impose certain restrictions on our behavior. An object of commitment can be the idea of fairness for example. This value is embodied in our norm of fair sharing. The commitment to this value requires conformity to the norm in order to implement our commitments in action. Commitment to fairness provides the agent with a reason for norm-obedience so to say. But this reason is not instrumental in nature, because the agent's judgment in such cases is not based on the principle of instrumental rationality. The agent's performance of the action, i.e. his conformity to the social norm is independent of his personal situation; it is independent from his self-interest and his preferences. Commitment provides practical reasons that are "independent of the gains and losses for the person in case he or she acts on that reason" (Pauer-Studer 2007, 75).³¹

5.2 Theoretical Approaches to Commitment

Very few economists are interested in conceptualizing commitment as a source of observable behaviour. Among them are Amartya Sen and Thomas Schelling. Each approaches the topic from slightly different angles and they do so for different purposes. A quick outline of their main picture might improve our understanding of what commitments can be besides the rather intuitive characterization given above.

Schelling wrote extensively about commitment and is even sometimes credited with having originated the concept. In the first essay *Strategies of Commitment* in his book named *Self-Commitment and other Essays* (2006), he uses commitment to refer to phenomena such as "becoming committed, bound, or obligated to some course of action or inaction or to some constraint on future action. It is relinquishing some options, eliminating some choices, surrendering some control over one's future behaviour. And it is doing so deliberately, with a purpose. The purpose is to influence someone else's choices. Commitment does so by affecting that other's expectations of the committed one's behavior." (Schelling 2006, 1) Thus, Schelling strongly relates commitment to solutions of cooperative problems. As commitments can easily be broken, devices are necessary to them credible (e.g. commitments such as threats or promises require devices) credible, to signal or communicate commitments when we engage in interaction and

³¹ Note that commitments can change over time in the sense that we change the objects of our commitments. A member of the German conservative party, being committed to the values, goals and ideals the party represents, switches to the Green party, because he changes his objective of commitment to what the green party stands for. However, as commitments seem to closely relate to what we commonly understand as personal convictions and that they play a big part in making up one's identity, it is assumed that these changes occur very rarely; objects of commitments and commitments themselves seem to be much more stable than our tastes and preferences. This is an important issue. However, its proper discussion would go beyond the scope of this essay.

bargaining.³² In Schelling's view, the concept matters for explaining situations where players have to rely on each other to achieve the maximum benefit (i.e. collective action problems). For example Schelling claims that a commitment sometimes lends itself to collective action when people make a common commitment in the case of a strike. "Being committed to share one another's strike burdens is intended to reduce the aggregate burden by enhancing the local's commitment, and probably does." (2006, 4) In this account, commitments matter for 'coordinating' social action to collectively achieve the maximum benefit and solve collective action problems.³³

Although the view depicted by Schelling covers certain aspects we ascribed to the motive of commitment, Amartya Sen comes closer to our characteristic of commitment. His writings help us to demarcate more clearly the concept of commitment from other motives. In his article *Rational Fools: A Critique of the Behavioral Foundations of Economic Theory* published in 1977, Sen famously argues against the reduction of all motives for action as ultimately leading to preference-satisfaction and maximization. He remarks that the problem with economic theory is that "there is no choice-independent way of understanding someone's attitude towards alternatives", because there is at least one source of action that does not involve choice in the usual sense, commitment (Sen 1977, 323).

To introduce his idea, Sen draws on a distinction between the motive of sympathy and a motive of commitment. Although being a 'pro-social' motive, sympathy does not require a departure from individual utility-maximization. Similar to acting from sympathy outlined by Kant in his *Groundwork of the Metaphysic of Morals* in 1785, an action motivated by sympathy is ultimately self-regarding, i.e. "the concern for others directly affects one's own welfare" (Sen 1977, 326). Kant introduced the friend of mankind, who finds pleasure in spreading joy around him and who can take delight in the satisfaction of others. But Kant identifies this kind of satisfaction as an inclination like any other desire a person can have. The agent performs the action because the action has a positive impact on his well-being; he does not perform the action because the spread of joy is in itself a valuable action. He concludes that such an action, even if admirable, has no moral worth (Kant [1785]1997); the agent did not have the proper

³² What is missing in Schelling's account, namely these commitment devices to enforce commitments and make them credible is given by Robert Frank (1988). He assigns this task to emotions as commitments cannot be reliably communicated. Emotions fulfil this job by being strategically useful commitment devices and signalling devices for communicating the adherence of the commitments by an opponent by facial expressions for example.

³³ Schelling also talks about self-commitment or self-command in the sense that an individual has multiple selves at different point in times. "There are two or more sets of values alternately replace each other; or an unchanging array of values is differentially accessible at different times [...] In common language, a person is not always his usual self; [...] it looks as if different selves took turns, each self wanting its own values to govern what the other self or selves will do by way of eating, drinking, getting tattooed, speaking its mind, or committing suicide." (2006, 71) Taking up this metaphor on different selves, we can only indicate a relation of Schelling's idea to our characterization of commitment in saying that one self is motivated by preference-satisfaction and maximization and the other self judges or ranks his options according to his values, as mentioned before.

motivation to make his action to be morally worthy. In contrast, an action which is motivated by duty has moral worth; it is characterized as an action performed independently of personal desires and its effect on personal well-being. Kant illustrates this point by describing a philanthropist acting kindly towards others. He does not have any inclination which would positively or negatively affect his state of affairs. He simply acts because duty demands it. Being motivated in this way makes the action genuinely morally worthy from Kant's point of view.

Sen argues that abstracting from one's personal desires and inclinations implies that the action is not driven by them. Rather, an action motivated by commitment requires a "counterpreferential choice" from the part of the agent (Sen 1977, 328).³⁴ This does not require, that the agent is not allowed to have any kind of additional inclination which makes the action attractive for him. This just means that under at least one counterfactual condition, the utility expected by the act chosen would be unaffected.³⁵ Hence, the agent would unconditionally perform the action, i.e. would have no choice in the usual sense. He would still perform the action, even if he expects negative utility and/or even if it is at the expense of an alternative action that would make him better off. Thus, taking commitment as a possible motive for action, the identity of personal welfare and individual preferences resulting in choice no longer holds (Sen 1977, 1985 and 2005).³⁶ His commitment makes performing the action necessary for him for other reasons than preference-satisfaction. That means even if these inclinations or desires would not be present, he would still perform the action in question. Personal welfare-improvements are not the *reason* for the action of the agent; they are not what the agent aims at in the first place (Sen 1977; Hausman 2005). The requirements of sticking to one's commitment rather present the agent with a reason to act. However, this reason is normative in character.³⁷

5.3 Commitments and Morality

It seems obvious that there is a close relationship between commitment and moral agency. According to Kant, acting from duty requires a sense of moral commitment. Being committed to moral principles is a necessary precondition

³⁴ In this sense, the motive of commitment is to be strictly distinguished from other, socially orientated, motivations such as altruism, sympathy, among others. Actions motivated by them are modelled as preferences of the agent and thereby having a positive effect on one's personal utility. They are motivated by preference-satisfaction. Thereby they result in the maximization of individual utility and consequently are being revealed by individual behavior.

³⁵ It has to be noticed that it is difficult to imagine different states of the world and potential choices for action. However, we will not discuss this difficulty further here but just present the simple idea of commitment.

³⁶ This identity mainly depends on the underlying understanding of the concept of a preference or the nature of reasons. As claimed before, personal welfare and individual utility are (technically) equalled in mainstream economics. This assumption provides the main basis for conclusions about what we understand as a preference and what the nature of reasons is (Sen 1977 and 1985).

³⁷ What is meant by a normative reason could be understood here as a reason which is directly provided by the norm. Here, the action is performed for its own sake. This is to be contrasted with an instrumental reason which is provided by the goal of the agent. Here, the action is 'used' as an instrument.

for acting from duty. Taking Kant's motive of duty as one 'subgroup' of the general idea of commitment, Sen's understanding of commitment is that an action motivated by commitment is *per definitionem* independent from the satisfaction of any kind of inclination or immediate desires the agent is driven by in a given moment. In the case of commitment, as is the case with actions done from duty, the agent considers the action as the *right* thing to do. Consequently, as when acting from duty, commitment requires the agent to be capable of undertaking judgment and evaluation, being self-reflected and conscious about his values and principles upon which he bases his evaluations.³⁸

However, although morality implies commitment, morality should not be equalled or confused with it. Commitment results in the *right* action, although this does not necessarily mean that it results in what is seen as *morally right*. A healthy lifestyle and ecological awareness can serve as examples: because a person commits himself to live a healthy life and to protect the environment, he considers it to be the right thing to do sports and to cultivate his own vegetables. He takes both actions seriously because they are demanded by her commitment to these objects, even if sometimes it is rainy and cold outside and he would prefer to stay at home. Equally, his preference can sometimes also dictate him to eat fast food. Then he is confronted with a conflict between his commitments and his preferences. And the right thing here is seen as what is dictated by his commitments. However, doing sports is not morally right. Hence, in the context of commitment, the right action is defined by the object of the commitment. A person may also be committed to an ideology, a political party, or a religious view. A person may be committed to an organisation, to a social contract, to (moral) principles or to god. All these can be objects one commits oneself to. *Right* in this context thus just means that the agent considers the action itself (or the outcome of the action) as worth pursuing, independent of the personal benefit he can expect from the outcome of the action. And the agent comes to know that it is *the right thing to do* because he evaluates and judges the action on the basis of the values he holds while comparing possible alternative actions. He does not draw on the evaluative principle of utility-maximization. In this sense, an action motivated by commitment is, as a moral action, unconditional and stable. But it does not necessarily aim at what is the morally right thing to do.

In sum: the idea of commitment comprises an evaluation of possible alternatives that is not necessarily carried out by basing it on moral values; instead, it can be based on any set of values the individual in question holds. However, what commitment has in common with morality, at least understood in a Kantian way, is the existence of a non-welfarist goal that is altogether removed from individual preference-satisfaction and the fulfilment of immediate inclination.

³⁸ The character of commitment strongly depends on the specific definition of concepts used and the specific 'character-traits' considered to be the most important one's for its final conceptualization. Frankfurt-minded philosophers would interpret commitment in a less evaluative and a less-cognitive way. According to Frankfurt people do not always know what they are committed to. Sometimes they believe they are committed to something they actually are not for example (e.g. Frankfurt 1971). We just use Sen's understanding of commitment here, which seems very much in line with a Kantian view of moral agency.

Although commitment as a motive for behaviour is not treated very extensively in the RC-literature, it is an important motive. It is important because it does result in behavioural patterns different from the ones resulting from individual preference-satisfaction. As already indicated, observed behavioural patterns will differ in stability of conformity depending on an individual's motivation. This matters in the case of organizational and incentive design. Workers who conform to the rules and norms prescribed by their companies out of a personal commitment to their company or to its goals and ideas, for which it stands, will conform to work norms to bring about these goals and will, at least to a certain extent, disregard personal interests. The same could be said for members and supporters of political parties. A person committed to the ideals and overall goals of a political party will agree more readily to the collective provision of certain public goods and thus will not attempt to free-ride; thereby, collective action problems can be avoided without incurring costs. On the contrary, a committed person might voluntarily punish people who consciously free-ride, even if this entails having to bear the costs that are incurred. It also matters in case we want to explain and predict behaviour. The behaviour of a committed person is less flexible, because it is not contingent upon circumstances or changes in preferences. The economist finds himself in an explanatory dilemma, as RCT does not allow for such a concept.

6. Implications

What would be the conceptual and methodological implications of a) acknowledging a normative dimension of agency which has several characteristics distinct from the instrumental dimension assumed by RCT; and of b) suggesting a conceptualization of the motive of commitment to accommodate the characteristic features of the normative dimension, which, however contradicts fundamentally with the behavioural assumption structure assumed in models based on RCT? We think that there are three ways to deal with this dilemma:

Firstly, many economists simply dismiss the motive of commitment as not being important, if at all, reducible to preferences. But this is problematic, because a concept of commitment seems to account for normative phenomena such as certain kinds of norm-conformity, work motivation, solidarity bonds etc. in which the normative dimension of agency is more defining for the behaviour in question (Pauer-Studer 2007) and it provides a useful account of cooperation observed in experimental settings (Kerr et al. 1997).³⁹ Additionally, as Sen himself notes: "if committed behavior is ruled out, then the reasoning that can go into the determination of choices would be correspondingly impaired. That would hardly be the way to give reason its due in the idea of rationality in general and rational choice in particular." (2005, 22ff.)

³⁹ We will not tackle the issue of providing empirical evidence for the existence of a motive of commitment in this paper. However, we are aware that this is a very delicate and important issue where a lot more work needs to be done. Interesting empirical literature is provided by Kerr et al. 1997.

A second possibility is to ignore the special nature of commitment and just assume that people have a preference for commitment; these approaches explain anomalies with as-if models based on long-term calculations and self-interest. This seems plausible at first sight. As mentioned in the first part of this article, economists have elaborated upon their concepts of self-interest and preferences thanks to the openness and flexibility of the utility-function. However, assuming a preference for commitment to explain committed behaviour leads to a tautological explanation, which provides little information. The *explanandum*, i.e. the event or fact to be explained is at the same time used as part of the *explanans*, i.e. the mechanism which does the explanation. Taking into account different types of (real) motives that people can have instead of reducing them to preferences, we could avoid basing our explanations on tautological statements and untestable hypotheses, as allowed by the revealed preference approach underlying RCT.

In addition, assuming a preference for commitment seems as paradoxical considering our everyday understanding of the concept; even modelling commitment as an ‘unconditional preference’ does not capture our intuition.⁴⁰ An action motivated by commitment does not imply consistent choice. We cannot infer from the agent’s choice that he has chosen the course of action with the highest utility. Such behaviour cannot be said to reveal our preference ranking because it is taken independently of how well it satisfies our preferences. It is a product of what our commitments demand. Hence, simply introducing a ‘taste for commitment’ fails to capture what is distinctive about the nature of commitment, i.e. that it can override considerations of preference-satisfaction.

The possibility of modelling commitment as an external constraint on the individual utility function also seems to be problematic in terms of capturing our commonsense understanding of the concept. People can break their commitments but they cannot remove external constraints such as budget constraints in mainstream consumer theory; they are externally given (Elster 1990). It is also questionable whether commitments restrict us in our maximization as they are self-imposed, or whether we should question the assumption of maximization altogether in this context.

Conceptualizing commitment as something having unconditional motivating power over any other preferences, however, also seems to be too strong a characterization. To make the concept more realistic and appealing to empirical research, we could investigate degrees of commitment within thresholds. So, when a starving person has committed himself to a healthy lifestyle and the idea of fair trade, he might nevertheless go to a fast-food restaurant, trading off his commitment against his immediate desire deriving from his hunger. But above a certain threshold (e.g. enough income to eat in the expensive whole-food restaurant)

⁴⁰ When we talk about an unconditional preference here, the concept is to be understood in the classical sense of a preference in economics, a binary relation, i.e. and “all-things-considered-ranking” (Hausman 2007; Bicchieri 2006), that satisfies the required conditions. In this case, the agent has a preference which is always ranked first in his ordering, no matter what other preferences he has. If I have an unconditional preference for coffee for breakfast, I will *always* strictly prefer coffee to *everything* else I could have for breakfast. This implies that having coffee for breakfast will always provide me with the greatest pleasure as it satisfies my unconditional preference for coffee. Hence, the choice of having coffee will maximize my utility.

he might trade off desire against commitments; the same holds for tastes. This threshold level can differ from individual to individual. However, to capture the general idea of commitment, we would have to assume that the threshold beyond which someone is willing to compromise commitments is relatively (noticeably) high.

A third possibility would be to take commitment seriously as a motivational force, to admit the explanatory shortcomings of the current framework and try to adapt it. In this case, the assumption of maximization and preference-satisfaction, i.e. the consistency-requirements for an instrumentally rationality choice would have to be relaxed or even rejected (Sen 1982). We defend the last option. To make use of the potential explanatory advantages of motives, we must improve the impoverished notion of preference, rethink the minimal requirements of consistency, face the challenge of our notion of reasons consisting of desires and beliefs that motivate behaviour, and change our understanding of what it means to act rationally.

7. Concluding Remarks

This article rests on the well-known observation that economists take preferences to describe behavior rather than people's real motivation. Based on the assumptions of revealed preference theory, it is assumed that an agent's motives can be inferred from observed behaviour in form of a preference for the observed behaviour that the agent is assumed to have. However, we have argued that real motives matter. They matter because experimental evidence has shown systematic deviations from the behavioural assumptions of maximization and (narrow) self-interest. People do in fact have a variety of different motives to perform specific behaviours, whereby behavioural patterns are conditioned on the nature of the motives involved and can differ accordingly. The first aim of the paper was to show that discerning different motives would provide us with a better understanding of why people act the way they do and thus to help us making better explanations and predictions of such behaviour.

Depending on people's motives, behavior can have more than one dimension. We aimed at showing here that at least an instrumental and a normative dimension exists. In the first case, people's tastes and preferences shape their motivation. In the second case, we argued that people's values shape their motivation. We took the example of norm-conformity: conformity to norms has an 'instrumental dimension', which is well captured by the economic framework and the notion of preferences. But it has also a 'normative dimension', which, so the argument goes, cannot be captured by RCT. Depending on people's motivation, they consider the conformity to social norms as normative either in one of the two senses or in both at the same time. We suggested the concept of commitment as a useful concept to capture this normative dimension of agency.

Motives also matter because experimental results do not tell us why people act in the way they do. Taking an 'as-if-interpretation' of RCT, any behaviour can be explained by referring to the agents' revealed preferences for that behaviour.

However, the actual causes of the behaviour and the mechanisms underlying it are requisite for arriving at useful explanations of behaviour. We are in need of more sophisticated hypotheses about what goes on in the black box, i.e. what really people motivates people. “To do this we have to pay attention to the meanings of the concepts we use (and test).” (Bicchieri 2006, 19ff.) To define our concepts and clearly demarcate them from others, it is useful to know why people act the way they do. Depending on their motivation, the reasons why people conform to norms will be different and thus the involved normativity, i.e. why people think they ought to follow the social norm, will differ. With it, the character of the performed behaviour will differ respectively in terms of long-term-stability, trade-off-considerations etc. By trying to outline what character traits a concept of commitment would need to reflect, a second aim of this paper was to show the difficulty economists have in accommodating the concept of commitment suggested by Amartya Sen.

To overcome the conceptual and explanatory shortcomings of the behavioural foundations of RCT, we have to consider motives as an important explanatory tool in social scientific modelling. It is widely recognized that people’s behaviour is complex. No theoretical framework will ever cover all of the potential types of motivations that exist. However, if RCT is to be taken as an explanatory enterprise and is expected to provide us with ‘reasonable explanations’ and predictions of people’s behaviour, its main assumption structure should have empirical content and should, at least to a large extent, represent our observations in a realistic way. While there is a rapidly increasing literature on motives such as altruism and reciprocity, there are unfortunately few empirical studies on the nature of commitment and the character of the behaviour a person motivated by commitment brings about. To extend the economic model, improve our utility functions and provide better cognitive and behavioural assumptions regarding the agents, the acknowledgement of the difference in the nature of motives and their irreducibility would be a step in the right direction. Conducting further theoretical and empirical research in this specific area is necessary for constructing a typology of motives. Being able to adequately conceptualize and formalize the fundamental differences in distinctively motivated behaviour is a first, but very important step towards improving social scientific models of human behaviour based on RCT.

Bibliography

- Anderson, E. (2000), Beyond Homo Economicus: New Developments in Theories of Social Norms, in: *Philosophy and Public Affairs* 29(2), 170–200
- Anscombe, G. E. M. (1957), *Intention*, Cambridge/MA
- Akerlof, G. A. (1982), Labor Contracts as Partial Gift Exchange, in: *Quarterly Journal of Economics* 97(4), 543–569
- Arnott, R. et al. (2003) (eds.), *Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz*, Cambridge/MA
- Arrow, K. (1971), *Essays in the Theory of Risk-Bearing*, Chicago
- Basu, K./J. W. Weibull (2003), Punctuality: A Cultural Trait as Equilibrium, in: Arnott et al. (2003), 163–182

- Ben-Ner, A./L. Putterman (1998) (eds.), *Economics, Values, and Organization*, Cambridge
- Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge
- Boland, L. A. (1981), On the Futility of Criticizing the Neoclassical Maximization Hypothesis, in: *American Economic Review* 71, 1031–1036
- Bunge, M. A. (1996), *Finding Philosophy in the Social Sciences*, New Haven
- Camerer, C. F./E. Fehr (2002), Measuring Norms and Preferences Using Experimental Games, Working Paper Series No. 97 of the Institute for Empirical Research in Economics/University of Zurich
- Coleman, J. (1990), *Foundations of Social Theory*, Cambridge/MA
- (1987), Norms as Social Capital, in: Radnitzky/Bernholz (1987), 133–155
- Conlin, M./M. Lynn/T. O'donoghue (2003), The Norm of Restaurant Tipping, in: *Journal of Economic Behavior and Organization* 52(3), 297–321
- Davis, J. (2004), Identity and Commitment, Tinbergen Institute Discussion Paper, 055/2
- Davidson, D. (2001), *Essays on Actions and Events*, Oxford
- (1963), Actions, Reasons, and Causes, in: Davidson (2001), 3–21
- Ellickson, R. C. (1994), *Order without Law: How Neighbors Settle Disputes*, Cambridge/MA
- Elster, J. (1998), Emotions and Economic Theory, in: *Journal of Economic Literature* 36(1), 47–74
- (1994), The Nature and Scope of Rational-Choice Explanation, in: Martin/McIntyre (1994), 311–322
- (1990), *Nuts and Bolts for the Social Sciences*, Cambridge
- (1989a), *Solomonic Judgements: Studies in the Limitations of Rationality*, Cambridge
- (1989), Social Norms and Economic Theory, in: *Journal of Economic Perspectives* 3(4), 99–117
- Fehr, E./U. Fischbacher (2004), Social Norms and Human Cooperation, in: *TREND in Cognitive Sciences* 8(4), 185–190
- / — (2002), Why Social Preferences Matter—The Impact of Non-selfish Motives on Competition, Cooperation, and Incentives, in: *The Economic Journal* 112, C1–C33
- /E. N. Kirchler/A. Weichbold/S. Gächter (1998), When Social Norms Overpower Competition, Gift Exchange in Experimental Markets, in: *Journal of Labor Economics* 16, 321–354
- Frank, R. (1988), *Passions within Reason: The Strategic Role of Emotions*, New York
- Frankfurt, H. (1971), Freedom of the Will and the Concept of a Person, in: *Journal of Philosophy* 68(1), 5–20
- Hardin, R. (2008), Norms and Games, in: *Philosophy of Science* 75, 843–849
- Hargreaves-Heap, S. (2009), Review of Bart Engelen's Rationality and Institutions: On the Normative Implications of Rational Choice Theory, in: *Erasmus Journal for Philosophy and Economics* 2(1), 132–138
- et al. (1992), *The Theory of Choice: A Critical Guide*, Cambridge/MA
- Hausman, D. (2005), Sympathy, Commitment, and Preference, in: *Economics and Philosophy* 21, 33–50
- /M. McPherson (2006), *Economic Analysis, Moral Philosophy, and Public Policy*, Cambridge
- Hechter, M./K.-D. Opp (2001) (eds.), *Social Norms*, New York

- / — (2001), What Have We Learned about the Emergence of Social Norms?, in: Hechter/Opp (2001), 394–417
- Heiner, R. A. (1983), The Origin of Predictable Behavior, in: *The American Economic Review* 73(4), 560–595
- Henrich, J. (2000), Does Culture Matter in Economic Behaviour? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon, in: *American Economic Review* 90(4), 973–979
- Hume, D. ([1739]1978), *A Treatise of Human Nature*, Oxford
- Jolls, C./C. Sunstein/R. Thaler (1998), A Behavioural Approach to Law and Economics, in: *Stanford Law Review* 50, 1471–1550
- Kahneman, D./J. L. Knetsch/R. Thaler (1986), Fairness as a Constraint on Profit Seeking: Entitlements in the Market, in: *American Economic Review* 76(4), 728–741
- Kant, I. ([1785]1997), *Groundwork of the Metaphysics of Morals*, Cambridge
- Kerr, N. L./J. Garst/D. A. Lewandowski/S. E. Harris (1997), That Still, Small Voice: Commitment to Cooperate as an Internalized versus a Social Norm, in: *Personality and Social Psychology Bulletin* 23, 1300–1311
- /C. Kaufman-Gilliland (1994), Communication, Commitment, and Cooperation in Social Dilemmas, in: *Journal of Personality and Social Psychology* 66, 319–332
- Kirchgässner, G. (2006), Das Ökonomische Verhaltensmodell: Der Homo Oeconomicus, in: Nell/Kufeld (2006), 81–107
- Klamer, A. (2003), A Pragmatic View on Values in Economics, in: *Journal of Economic Methodology* 10(2), 1–24
- Krupka, E./R. Weber (2008), Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?, in: *Discussion Paper Series Institute for the Study of Labor (IZA) DP No. 3860*, URL: http://www.iza.org/index.html?lang=en&mainframe=http%3A//www.iza.org/en/webcontent/personnel/photos/index_html%3Fkey%3D3557&topSelect=personnel&subSelect=staff [6.12.2008]
- Lindbeck, A./S. Nyberg/J. W. Weibull (1999), Social Norms and Economic Incentives in the Welfare State, in: *Quarterly Journal of Economics* 114(1), 1–35
- Mansbridge, J. (1998), Starting with Nothing: On the Impossibility of Grounding Norms Solely in Self Interest, in: Ben-Ner/Putterman (1998), 151–168
- Mantzavinos, C. (2009) (ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, Cambridge
- (2001), *Individuals, Institutions, and Markets*, Cambridge
- Martin, M./L. McIntyre (1994) (eds.), *Readings in the Philosophy of the Social Science*, Cambridge
- Millar, A. (unpublished), The Explanatory Role of Normative Considerations, URL: <http://www.essex.ac.uk/philosophy/tpn/documents/Normativity/MillarExplanatoryRole.pdf> [10.6.2006].
- Nell, V./K. Kufeld (2006) (eds.), *Homo Oeconomicus: Ein neues Leitbild in der globalisierten Welt?*, Berlin
- Von Neumann, J./O. Morgenstern (1944), *Theory of Games and Economic Behavior*, New Jersey
- Okruch, S. (1999), Innovation und Diffusion von Normen: Grundlagen und Elemente einer evolutorischen Theorie des Institutionenwandels, in: *Volkswirtschaftliche Schriften Bd. 491*, Berlin
- Opp, K.-D. (1999), Contending Conceptions of the Theory of Rational Action, in: *Journal of Theoretical Politics* 11(2), 171–202
- Ostrom, E. (2000), Collective Action and the Evolution of Social Norms, in: *Journal of Economic Perspectives* 14(3), 137–158

- Pauer-Studer, H. (2007), Instrumental Rationality versus Practical Reason: Desires, Ends, and Commitment, in: Peter/Schmid (2007), 73–104
- Peter, F./H. B. Schmid (2007) (eds.), *Rationality and Commitment*, Oxford
- Pettit, P. (1990), The Reality of Rule-Following, in: *Mind* 99(393), 1–21
- Radnitzky, G./P. Bernholz (1987), *Economic Imperialism: The Economic Method Applied Outside the Field of Economics*, New York
- Raz, J. (2004), The Myth of Instrumental Rationality, URL: <http://www.jesp.org/PDF/6863Raz-vol-1-no-1-rev.pdf> [21.8.2007]
- Rosenberg, A. (1995), *Philosophy of Social Science*, Boulder
- Samuelson, P. A. (1938), A Note on the Pure Theory of Consumer's Behavior, in: *Economica* 5(17), 61–71
- Schelling, T. (2006), *Strategies of Commitment and Other Essays*, Cambridge/MA
- Schmid, H. B. (2007), Beyond Self-Goal Choice: Amartya Sen's Analysis of the Structure of Commitment and the Role of Shared Desires, in: Peter/Schmid (2007), 211–226
- Searle, J. (2001), *Rationality in Action*, Cambridge/MA
- Sen, A. (2005), Why Exactly is Commitment Important for Rationality?, in: *Economics and Philosophy* 21, 5–13; reprinted in: Peter/Schmid (2007), 17–27
- (2002), *General Introduction to 'Rationality and Freedom'*, Cambridge/MA
- (1987), *On Ethics and Economics*, Cambridge/MA
- (1985), Goals, Commitment, and Identity, in: *Journal of Law, Economics and Organization* 1(2), 341–355
- (1982), *Choice, Welfare and Measurement*, Cambridge/MA
- (1977), Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, in: *Philosophy and Public Affairs* 6(4), 314–344
- Stout, L. A. (2001), Other-regarding Preferences and Social Norms, Working Paper Series in Business, Economics and Regulatory Policy and Public Law and Legal Theory, No. 265902, Georgetown University Law Center, URL: <http://papers.ssrn.com/sol3/papers.cfm?abstractid=265902> [2.5.2008]
- Smith, A. ([1795]1976), *The Theory of Moral Sentiments*, eds. D. Raphael and A. Macfie, Oxford
- Sugden, R. (2000), Team Preferences, in: *Economics and Philosophy* 16, 175–204
- Turnbull, C. (1972), *The Mountain People*, New York
- Wallace, R. J. et al. (2004) (eds.), *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, Oxford
- Woodward, J. (2009), Why Do People Cooperate as Much as They Do, in: Mantzavinos (2009), 295–357
- Zimbardo, P./R. Gerrig (1996), *Psychologie*, Springer