

Los trabajos de C. Castrodeza, de entre los cuales este libro supone la culminación, constituyen una referencia singular e indiscutible en un panorama intelectual como el español muy poco atento a la sensibilidad científica naturalista. “La darwinización del mundo” es un texto valiente y atrevido, que hace pensar y ofrece nuevas claves para comprender la realidad humana sin renunciar ni a la radicalidad del pensamiento filosófico ni al rigor científico.

Miguel A. Castro Nogueira  
& Laureano Castro Nogueira  
UNED  
lcastro@madrid.uned.es

FABIENNE PETER & HANS BERNHARD SCHMID, eds. 2007. *Rationality and Commitment*. Oxford: Oxford University Press.

Amartya Sen's “Rational Fools” (1977) is nowadays one of the most cited and commented works in the field of rational choice theory (RCT). Although Sen has developed his views in greater depth in subsequent works, his initial claim has not lost its critical power.

From Sen's point of view, commitment cannot be accommodated in RCT explanations because it opens a wedge between welfare and choice. In “Rational Fools”, Sen argued that we must distinguish between two separate concepts: sympathy and commitment. The former corresponds to the case in which the concern for others directly affects one's own welfare: “If the knowledge of tortures of others make you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment” (Sen 1977, 319). Later, in “Goals, Commitment and Identity” (1985), Sen developed the theoretical distinction between self-centered welfare, self-welfare goal, and self-goal choice, and placed this distinction in the core of RCT models. He argues that sympathy only violates the self-centered welfare condition, because the welfare of others influences our own welfare. RCT can easily explain this kind of “altruism”, due to the fact that an agent's welfare increases by making other's welfare increase as well. Sen argues that commitment involves making a choice which violates either the RCT requirement of self-welfare goal or self-goal choice. Sen claims that “commitment is concerned with breaking the tight link between individual welfare (with or without sympathy) and the choice of action (for example, being committed to help remove some misery even though one personally does not suffer from it)” (Sen 1985, 7-8).

*Rationality and Commitment* is the result of a workshop that was held in the University of St. Gallen in May 2004. It collects a series of essays that aim to discuss Sen's critique of the implicit selfishness of a narrow interpretation of rational agency, under the light of the recent development that RCT has experienced, as well as the debates on rationality and collective agency than have taken place in philosophy.

This volume is divided in three sections. Under the heading 'Committed Action', the first three chapters are devoted to explore the relation between commitment, choice and preference, starting from Sen's review of his concept of commitment (Chapter 1). Following Sen, sympathy would violate self-centered welfare (due to the fact that the agent's welfare depends on the welfare of others), whereas commitment can imply modifying the agent's goals in a way that she does not pursue her own welfare (therefore violating the self-welfare goal principle), as well as altering the agent's choice by incorporating other person's goals (therefore violating the self-goal choice). Sen argues that commitment is important to understand and rethink rationality for two reasons. First, rationality should take note of conduct that may not be congruent with the pursuit of one's own goals and one's own welfare, no matter how broadly characterized. Second, commitment can help to explain many patterns of behaviour that we actually observe and that are hard to fit into the narrow format of RCT. In Chapter 2, Pettit focus on Sen's characterization of commitment as a violation of the self-goal choice assumption. Drawing a distinction between goal modifying and goal displacing, Pettit argues that, while the modification of an agent's goals in order to consider other people's goals is quite common, the possibility of acting in order to attain a goal that the agent does not have is highly implausible: the notion of agency entails a relation between an agent's goals and actions. While goal displacing requires a departure from RCT, goal modifying can be accommodated by modelling the deliberation process that allows agents to include other agent's goals as owns. Hausman (Chapter 3), too, believes that commitment can be accommodated within the RCT framework. If preferences are seen as all-things-considered rankings, then commitment does not entail counter-preferential choice. Rather, Hausman argues, commitment should be invoked as one of the preference formation mechanisms, this is, as a kind of motivation, among others.

'Rethinking Rationality' is the second part of the book, and includes a collection of reflections around the relationship between rationality and commitment, both within and from outside RCT. Pauer-Studer (Chapter 4) compares humane versus kantian accounts of rationality, and argues that Sen would fall under the category of a moderate kantian. She argues that a broader conception of instrumental rationality that does not entail self-interest rationality, and can be compatible with a formal model such as RCT. In Chapter 5, Brennan makes a distinction between two accounts of rational agency: as an 'approach' and as a theory. The rational agent account does not make any assumption about what the content of the agent's preferences is, but rather explores the formal restrictions on the structure of preferences. Then, through the analysis of two of Sen's examples (the voting problem and the boys and the apples example of committed action) Brennan concludes that the characterization of commitment as a counter-preferential choice is misleading. From a similar point of view, Güth and Kliemt (Chapter 6) also advocate for a distinction between a formal framework (what they call 'Rational Choice Modelling') and a theory of rationality, as the RCT (which would include some assumptions about the content of the agent's preferences). They argue that commitments can be accommodated in Rational Choice Modelling, by modifying the rules in the game, either through a payoff modification,

or through the elimination or addition of moves. Verbeek's contribution (Chapter 7) brings together Sen's approach and Elster's theory of pre-commitments, and discusses the intertemporal dimension of rationality and choice. Verbeek argues that commitment has both a causal and a normative dimension. Elster's pre-commitment devices consist in raising the cost of performing an action in the future, but from his point of view, making a decision is costless. Verbeek argues against this claim: choosing a path of action provides the agent a new reason for acting as decided, thus making commitments rational, independently of the payoff structure at the time of performing the action. In the eighth and last chapter of this section, Gächter and Thöni overview the empirical research in cooperation problems such as the Prisoner's Dilemma and the public goods game, which reflect a much higher rate of cooperation than theoretically predicted. The observed results, they argue, do not contradict the RCT model, but they challenge the universal selfishness assumption.

In the third part of the book ('Commitment, Intentions and Identity'), we find a series of contributions which explore the relationship between commitment and collective intentionality. Schmid (Chapter 9) argues that, in order to include commitment in the explanation of action, it is not enough to widen our concept of motivation: reasons, and not motivations, are the basis of committed action, which has to do with shared goals, holding a normative relation with individual contributive goals. Shared goals are the source of reasons for action, which are neither internal nor external, in Williams' terminology, but "transcend one's subjective motivational set" (p. 222). In Chapter 10, Tuomela develops some aspects of his theory of social action, based on the distinction between the I-mode and the we-mode perspectives. He argues that standard game theory is concerned about I-mode cooperation, and cannot capture the kind of cooperation between individuals from the we-mode. Commitment inside groups, Tuomela argues, can be explained by analyzing the group's preferences and goals, and how individuals contribute to the promotion of the group's view, its Ethos. From a similar perspective, Gilbert (Chapter 11) focuses on the concept of joint commitment, which she has developed in several previous works. She claims that collective action has an intrinsic normative dimension: agents commit their wills to perform their part in a joint action, binding themselves through the recognition of their agreement (whether implicit or explicit) as a "sufficient reason to act in accordance with the commitment" (p. 267). Gold and Sudgen (Chapter 12) review different approaches to team reasoning, which attempt to explain two puzzles of game-theoretical explanations: the Prisoner's Dilemma and the Hi-Lo game. Theories that are based on an individual account of agency may use payoff transformation, or preference patterns (instead of unique preference rankings) in order to explain the experimental results in these games; Gold and Sudgen, however, argue that focusing on team (rather than individual) agency allows to introduce team reasoning, which performs better in explaining these puzzles. In the last Chapter of the book, Davis explores the implications of Sen's critique to the problem of the identity of the agents. He analyses Sen's views on social identification, and review Bernard Williams' concept of commitment as identity-conferring. He argues that the reflexive dimension of commitment, related to the capacity of reasoned self-scrutiny, is constitutive of the agents' identity.

The final section of the book is a comment by Amartya Sen on the contributions collected in the volume. He discusses the views of the different contributors, focusing his answers on Pettit's and Hausman's chapters.

Overall, *Rationality and Commitment* collects much of the recent debate around the challenge posed by Sen more than three decades ago: what the relevance of committed action is in the explanations of behavior, and how to reconcile commitment and rationality in order to provide a more realistic picture of human motivation and action. The contributions in this book, including Sen's introduction and concluding remarks, have the capacity of summing up the current state of a problem that is still present both in philosophy and economics, allowing the reader to approach a controversy that is far from being settled. Nonetheless, sometimes the reader might think that some of the topics this book covers are weakly interconnected, or somehow independent. This is due to the difficulty in dealing with problems that are interdisciplinary in nature, and also to the fact that the concept of commitment is used in very different contexts, as Gilbert points out (p. 261). Most of the essays in this book discuss the different implications of Sen's concept of commitment, but some of them broaden the term, especially those concerning collective intentionality. However, I think that the ambiguity of the concept of commitment reflects the present state of the discussion, and not considering other uses of the concept would limit the scope of this volume. In brief, it is not only a review of Sen's critique, but also a useful introduction to a problem that is central to the explanations of human agency.

Miranda del Corral  
UNED  
miranda.delcorral@gmail.com