Rational Choice and Collective Decisions

Rational Choice and Collective Decisions:

A Primer

By

Diego Lanzi

Cambridge Scholars Publishing

Rational Choice and Collective Decisions: A Primer

By Diego Lanzi

This book first published 2025

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-4289-7
ISBN (Ebook): 978-1-0364-4290-3

A man has free choice to the extent that he is rational.
—*Thomas Aquinas*

TABLE OF CONTENTS

PREFACE

In his "*Rational Fools: A Critique of the Behavioral Foundations of Economic Theory*", published in 1977, Amartya Sen stresses the following:

> I have tried to analyze the structural extensions in the conception of preference made necessary by behavior based on reasoned assessment of commitment. Preferences as rankings have to be replaced by a richer structure involving meta-rankings and related concepts (p. 344).

Sen argues that the main issue is the acceptability of the invariable pursuit of self-interest in each rational act. Considering that type of behavior rational and departures from it irrational yields an arbitrarily narrow definition of rationality and depicts individuals as rational fools unable to assume commitment, feel emotions and elicit values and virtues other than utility maximization.

In particular, Sen is interested in the role of moral and ethical commitment that, in his own words,

> does not presuppose reasoning, but it does not exclude it; in fact, insofar as consequences on others have to be more clearly understood and assessed in terms of one's values and instincts, the scope for reasoning may well expand (p. 334).

Surely, any theory of rational choice presupposes that choices are influenced by reason. This not only implies that we are reasoning beings who possess cognitive abilities and apply principles of logic (deduction/induction) to understand the world but also that we use reasoning to guide our behavior/decisions/choices. The scope of such reasoning, however, is the central issue.

There are two opposite ways to delimit the scope of rational reasoning.

On the one hand, we have the concept of *instrumental rationality*, suggested by David Hume. According to this view, reason alone cannot be a motive of action since it is the slave of passions. Human choices are ultimately driven by passions, desires and wants, and the rationality of an action consists of optimally using all available instruments to achieve

desired goals. Passions or desires themselves cannot be irrational but can be based on false or deceptive judgments. Conversely, irrationality is the choice of wrong means to pursue some ends.

On the other hand, Immanuel Kant suggested the idea of *principled* rationality. Moreover, Kant was persuaded that passions and desires were alien influences that should not affect human reasoning. From a Kantian perspective, if you desire to achieve *y*, and this can only be accomplished by the course of action *x*, then it is for you instrumentally, or hypothetically, imperative to do *x*. However, in a well-ordered society, human behavior must be ruled by *categorical imperatives*, i.e., moral dictates recognized by individuals who possess the faculty of autonomously acting according to the maxim: "*you can that it should become a universal law of behavior*". In this sense, Kantian rationality is fueled, or principled, by morality, sobriety and integrity.

Since the beginning of the last century, many refinements and intermediate conceptions of rational behavior have been proposed. Surely, one of the most notable proposals has been the idea of *bounded*, or procedural, rationality suggested by Herbert Simon (Simon (1955), (1972), (1982)).[1] As Simon (2000) noted,

> Human beings are adaptive organisms and social organisms that can preserve body of learning or can be influenced by socially, culturally, or positionally determined norms, values, or habits, studying their behavior will not return us to permanently invariant laws, for human learning, and social influence will continue to change people's way of making rational decisions (p. 252).

Simon's research has been focused on information- or knowledge-related limits to rationality. He has suggested a procedural view of rational behavior, specifying how imperfectly rational choices vary under different choice procedures and different notions of what a *satisfying outcome* is. Furthermore, Elster (1979; 1983) analyzed other cases of limited rationality due to, for instance, weakness of will or adaptation of aspirations to what is actually feasible. Indeed, a third way for the rationality concept is possible.

However, both Homer's Ulysses, with his tying-the-hands strategy, and Pedro's fox do not avoid foolishness. The first denies his desire to hear the

[1] An elegant step forward in modelling procedural rationality has been made by Rubinstein (1998).

voice of Sirens and to follow their sensual invitation. The second fills in its mind with wishful thinking. They deliberately ignore their emotions and real desires and capabilities; they fail to recognize their virtues and vices and do not change their way of thinking because of contextual conditions. Ulysses remains who Seneca describes, in his tragedy *Troades*, as "*machinator fraudis et scelerum artifex"*,[2] the fox convinced of the false.

[2] Translation: the engineer of fraud and the craftsman of crimes.

INTRODUCTION

Two alternative approaches, albeit not reciprocally exclusive, to the analysis of individual choice behavior exist. A *normative approach*, i.e., the study of how individuals should behave if they were perfectly rational (that is, what should their behavior be according to a normative notion of rationality), and a *behavioral approach*, aimed at providing explanations of effective decisions, observed and adopted deliberately by non-perfectly rational individuals. Normative analysis of choice focuses on what an agent should do to be rational, following well-defined principles, properties and axioms. Hence, normative and behavioral choices often do not coincide. This happens because, in real-world decisions, agents do not follow lines of action prescribed by perfect rationality but rather reasonableness, intuition, or conduct dictated by limited and imperfect rationality. What usually makes normative results so divergent from behavioral results is the *embeddedness of choice*. Individuals, whenever they make a choice, find themselves embedded in a precise institutional setting composed of several organizational, political or social contexts. Any context has certain influences on individual choice behavior that can make preferable a non-perfectly rational action to conduct conforming to precepts of full rationality. Nevertheless, normative analysis is useful for understanding the logic of perfectly rational choices, knowing that this is uniquely an approximation of real choice behavior. The latter tends to converge toward what is enshrined by normative choice theory with well-informed agents with perfect computational abilities, competences and skills.

James Buchanan has proposed several criteria for categorizing normative decisions assumed by rational agents: the *incidence* of decisions, i.e., whether choices directly affect an agent's welfare, and *responsibility* over outcomes, i.e., how much an agent can control, and has the responsibility for, choice outcomes. By crossing these two criteria, he obtains a simple map of *rational choices*.

When individual choices have a strong and direct effect on an agent's welfare and these decisions yield outcomes directly imputable to the agent's responsibility, we have *individual choices*. If the responsibility for the final outcome is high but the choice effects on the agent's welfare are

weak, we have choices within *agency relations*. In contrast, if the incidence of decisions is high but each agent has low control over final outcomes, the choice problem is no longer individual but rather somewhat strategic. In this case, we have *strategic choices* analyzed by game theory. Finally, we have *collective choices* in which individual responsibility over outcomes is almost null, and the incidence of decisions on individual welfare is quite low. In this textbook, we focus on the first and last domains in the case of no uncertainty over choice alternatives.[3]

The basic normative notion we endorse is rationality. The paradigm of rationality commonly used in choice theory is the idea of instrumental rationality originally proposed by David Hume. An agent, following Hume's thought, is rational when it chooses efficient means for achieving some given ends. As we have seen, a notion of rational behavior is called *instrumental* because it relates solely to means/instruments to reach some ends, not discussing how these goals are established by values and passions.

Instrumental rationality has a dual nature depending on whether individual aims are multiple. Take an agent who has one single objective, for example, climbing the highest mountain in Europe. In this case, the idea of instrumental rationality says that he/she will be rational by trying to climb Mont Blanc. Instrumental rationality in these cases assumes traits of *rational satisfaction* (Harsanyi (1977)). Instead, if an agent has multiple goals, he will be rational in selecting the best outcome in terms of *preference satisfaction*. In this case, the comparison of different choice alternatives, once inaccessible options are removed from the choice set, takes place through an order relation named *individual preference*.

This book provides an introduction to the basics of individual and collective choice theory. It aims to deliver a survey of the main contributions of the discipline for last-year undergraduate students, or Ph.D. students. For simplicity, we omit formal proofs, which can be found in the original references. However, for the sake of clarity, we have otherwise formally stated definitions, axioms and propositions.

The essay is organized in three chapters that are readable even separately: the first is devoted to discuss the notions of rationality and rational

[3] Regarding strategic choices, see any text of game theory, for example Binmore (2007) or Osborne and Rubinstein (1994); on expected utility theory and choices in conditions of uncertainty see Machina (1987) and Kreps (1988), whilst on agency models see Hart and Holmstrom (1987).

behavior; the second reserved to individual choices, and the last one focused on the problem of social choice. Examples, which explain main presented propositions, might also be used as exercises in classrooms. In the Appendix, other review exercises are proposed.

In Chapter 1, we discuss the concept of rationality, its domains and its main interpretations. We analyze what the paradigm of rationality in sciences is and which of its versions is embraced in this essay. A multidisciplinary approach is used in order to clarify to the reader the focus of the book and its methodological assumptions.

In Chapter 2, we start from the very basic notions of choice and preference and discuss what a choice structure is and how it can be represented. Then, we move to more specific topics, such as preference revelation, consistency of choice, menu independence or decisiveness; hence, we drop some assumptions on preference relations and analyze incomplete and intransitive preferences and their effects on choice behavior. In the last part of the section, we briefly review the main critics to individual choice theory and several of its possible extensions.

In Chapter 3, we introduce the problem of aggregating individual preference orderings and the notion of collective preference. The chapter opens with voting theory and presents, initially, the main voting rules and their properties/drawbacks. Then, social decision rules and social choice functions are presented and discussed. In this way, we get tools and categories to consider the main theorems related to the impossibility/possibility of social choice. Several paragraphs devoted to social welfare functions, their invariance, and utility interpersonal comparability end the section.

Finally, in the conclusion, we bring the case for irrationality to light. We discuss main reasons of irrational decisions and main possible behavioral departures from the normative prescriptions of rational choice theory. In doing this, we open rooms for emotions, illusions, false beliefs, weakness of will, memory losses etc., all categories which are nowadays on the frontier of the academic research on rationality.

To conclude, we provide a list of classic surveys that can be used together with this book. Some of them are quite technical, whereas others are more history oriented. The interested reader can use them both to complement the book's contents and to obtain inspiration for further deepening of what is discussed here.

CHAPTER ONE

INDIVIDUAL AND COLLECTIVE: RATIONALITY AND CHOICE

Many academic disciplines address the issue of how individuals should behave according to the normative notion of *rationality*. We can consider them as the branches of the same fundamental science, what we can call a *theory of rational behavior*. But, what does rational behavior mean? When do we act rationally in making decisions? What is the hiatus between individual and collective rationality? In order to give answer to these, and similar, questions, a theory of rational behavior is needed.

Hence, the concept of rationality, its domains and its interpretations are object of this chapter. We aim to specify what the paradigm of rationality in sciences is and which of its versions is embraced in this essay. In doing this, we introduce several notions and concepts that we will find later in the book; just to quote some of them: freedom, democracy, utility, preference, commitment.

In what follows, we use a multidisciplinary approach to the notion of rationality for specifying the focus of the book and its methodological assumptions.

The Rationality Principle

The concept of rationality plays a crucial role in several social sciences from economics to sociology, politics, and organization theory and so on. The reason for such a centrality of rational behavior is twofold: on the one hand, rationality is a normative *desideratum*; on the other hand, human behavior in many empirical circumstances takes the shape of rational behavior. The rational approach to problem-solving is grounded in the view that, in order to explain individual or collective behavior, we have to use a model (or theory) of *rational choice*. A model of this sort examines whether an agent, individual or collective, has (or had) "good reasons" to behave as he/she does (did) in front of a given situation, or choice

problem. In other words, the *explicatum* of the idea of "rational action" is the "principle of rationality".

According to the principle of rationality, an action, or a choice, is considered as *rational* if and only if it satisfies the following six conditions.

(i) The agent *i* assesses the situation/problem as belonging at the class, or type, *S*;
(ii) The agent wants to achieve the goal *G*;
(iii) Given the situation *S*, he/she thinks that *M* is an effective and efficient mean for achieving *G*;
(iv) The agent has considered all benefits and costs (direct and indirect) of using *M* and he/she thinks that the benefits exceed the costs; in other words, choosing *M* is better than not to choose it;
(v) The agent has the abilities/capacities of using/choosing the mean *M*;
(vi) No constraints, individual or collective, forbid the use/choice of *M*, and then agent *i* uses/selects it.

In short, if above conditions hold, agent *i* has "good reasons" to use, or select, *M*. If he/she does not, *eo ipso* is behaving irrationally.

Rationality and Methodological Individualism

A large majority of models, or theories, of rational behavior embraces *methodological individualism (*MI*)*. The doctrine of MI suggests reducing all collective, or social, phenomena to the actions, interactions, goals, thoughts of individuals, and to the norms, or institutions, created by them. MI was introduced at the beginning of the XX century, by the works of Ludwing Von Mises, Friedrich Hayek and Karl Popper, to contrast the widespread use of holistic and communitarian paradigms.[4] The term was coined by Popper in his *The poverty of historicism*, published in 1945. From the advent of the Vienna school until today, such methodological perspective has been largely used by both economic and sociological theories of rational behavior.

In sociology, MI is the very basic principle of analysis in the *Weberian approach* to social phenomena.[5] Sociologists as Weber, Pareto, Simmel or

[4] For details see Laurent (1994).
[5] See Weber (2013).

Trevor-Roper insist on the fact that the main aim of sociology is to explain "the social" as the result, or product, of individual actions. Therefore, it is crucial to understand (*verstehen*) the motivations of individuals and to view their actions as adaptive responses to the situation they face. Finally, it is necessary to seize the aggregate (social) consequences of these individual actions. The relevance of this aggregation step is paramount. Only by aggregating in a correct way well-explained individual actions, or choices, the *substantive* truth of social occurrences will be discovered.

According to the Weberian approach to sociology, individual actions are rational by definition. Following the *verstehen* principle, they are rational reactions to the situation faced by the agent. Thus, any individual and *intentional* course of action cannot be judged as irrational. Irrationality belongs to the realm of the unintentional. Nevertheless, the kind of rationality used by agents depends on, following Weber, the *structure* of the situation at stake. In some cases, individual behavior can be explained through classic notions of rationality; in some others, more sophisticated accounts of rational behavior are needed.

Notions of Rational Behavior

As we have already discussed, two diametrically opposed notions of rationality lie at the extremes of a spectrum of conceptions of what rational behavior means. On the one hand, we have *Kantian rationality*; on the other hand, we have *Humian rationality*. As emphasized in the Preface, in this book we address the case of *instrumental* (Humian) rationality.

From a Humian point of view, morality has only an instrumental value and not an intrinsic one (like according to Kant). Therefore, sometimes to follow moral dictates can be virtuous, some other times it cannot. However, morality is the slave of passions; perhaps it defines what goals we pursue, perhaps not. In any case, morality cannot be used to find optimal means for achieving human goals (exactly the opposite conclusion is reached by Kantian rationality; e.g., moral actions are the only rational ways to behave). Within the paradigm of instrumentally-rational behavior, we can distinguish between *substantive rationality* and *procedural rationality*.[6]

The idea of substantive rationality is fairly common in marginalist economics. The economic agent has a unique and superior goal: the maximization of

[6] The distinction was introduced in Simon (1976).

X, where *X* can be utility, profit, expected utility, hedonic satisfaction and the like. Usually, in decisional models of this school of thought, agents have perfect and complete information and perfect "problem-solving" abilities, two conditions which allow them to take *optimal* decisions.

In contrast, Simon's idea of procedural, or *bounded*, rationality rejects such "Olympic view" of rational behavior in favor of a less demanding, and more realistic, idea of *satisfying behavior*. Roughly speaking, do not pursue the optimal decision, the first best, if it is too costly or impossible, but settle for a "good decision", perhaps the second best. From this standpoint, what matters is finding reasonable and consistent decisional procedures, an array of *rules of thumb*, and defining what satisfying means. As we will see in what follows, the mathematical framework of theories of substantive rationality can be easily adapted to address procedurally-rational behavior.

Finally, we have the case of *expressive rationality*.[7] A general, and well-established, definition of expressive rational behavior does not exist. The idea of expressive rationality is largely used in much anthropological investigation, and it is concerned with what anthropologists typically find distinctive about human activity: the shared meanings people attach to what they do. When a behavior conforms to shared norms of a group:

> The judgment about what is worthy in an action is shared by others who subscribe to the norm. So, when one's action conforms (or breaches) a norm it is regarded as worthy (or shameful) for the same reasons by a group. This endows action with a symbolic dimension: the action becomes a voice to the shared reasons that find it worthy, so enabling people to say things about themselves to others who share that norm (Heap (2001), p.127).

Consistently, a behavior is expressively rational when it expresses something about the individual. Both instrumental and non-instrumental accounts of expressive rationality can exist.[8] In the first case, the expressive action is seen as a decision which gives agents indirect utility and hence determines, before the choice happens, individual preference orderings. In the second one, instrumental justifications for action cannot be evoked to explain behavior, and individual choices have to be assessed consistently with commitments, loyalties or norms that an agent cares about and which shape his/her identity. In this last case, expressive choices

[7] On expressive rationality see Heap (2001).
[8] See Engelen (2006).

are taken for their own intrinsic reasons rather than for achieving some self-welfare goals; in John Elster's words: "the action itself, rather than the outcome it can be expected to produce, is what matters".[9]

Rationality and Deductivism

Theories and models of rational behavior make extensive use of the *deductive method*. Firstly axioms, properties, procedures are carefully defined, and then their consequences are logically deduced. This propensity for deductive reasoning is consistent with the view that rational behavior can be predicted by using logics (from *logós*) and that *mathémata* (i.e., mathematical categories) allow us to substantially and truthfully understand it. Nevertheless, even with regard to deductivism, rationality theories differ widely.[10]

Take the case you have a family of axioms, or properties, say $a_1, a_2, \dots, a_n$, and that these, when combined one with the others, imply a consequence, *C*, which is manifestly false. How to react to such evidence? What can be, or should be, deduced from the falseness of *C*?

Once again, we have to opposite extremes. On the one hand, there is the *negationist* position. On the other hand, we find the *eliminationist* position. According to the former, the implausibility of *C* is only apparent since *C* is logically-consistent with the axioms which are indisputable. Hence, more investigations are needed on the very nature of *C* in order to understand why it reveals inconsistencies and therefore, rationalize it (Watkins calls this epistemological approach *modus ponens*). According to the latter, axioms $a_1, a_2, \dots, a_n$ have to be modified because only apparently right, since they predict an empirically false outcome. The issue becomes, in this last case, how to identify and define more reasonable axioms or properties (e.g., *modus tollens*). As the reader of this primer will appreciate, these two epistemological paths have played a crucial role in the development of both individual and social choice theory.

---

[9] See Elter (1986). We will not discuss in this book how to extend rationality theory's framework in order to deal with expressive rationality. I have tried to address the issue elsewhere. See Lanzi (2022b) for details.

[10] This point was emphasized, for the first time in the literature, by Watkins (1979).

A Theory of Rational Behavior

By combining different branches of scientific research, we can profile a *theory of rational behavior*. Some disciplines address the problem of individual choice behavior; some others analyze how agents behave inside social settings.

In principle, an individual agent taken in isolation faces choices in conditions of *certainty*, *risk* or *uncertainty*. In the first case, the chooser exactly predicts which outcomes are determined by his/her actions. In the case of risky decisions, the agent has to cope with the fact that his/her actions can yield different outcomes, but, at least, he/she knows the objective probabilities of these results. Finally, we have choice in conditions of uncertainty when the above probabilities are partially or totally unknown to the agent. According to tradition, the theory of rational behavior in conditions of certainty is named *rational choice theory*; instead, theories and models of rational behavior in conditions of risk and uncertainty compose a normative theory usually named *decision theory*. In chapter two, we will focus on the theory of rational behavior in conditions of certainty.

With regard to disciplines which analyze individual rational behavior inside social settings, we have to mention two rather different fields of study: *game theory* and *ethics*. In game theoretic models of rational behavior, two, or more, agents with conflicting interests have to find optimal strategies for pursuing their individual goals. In some social environments, rational agents can use cooperative strategies; in others, they might be force to use un-cooperative strategies. Differently, ethics addresses the issue of how rational individuals can find adequate rules of behavior to cooperate in achieving common goals of the whole society.

In fact, all disciplines listed until now, analyze the rationality of individual actions. In contrast, *social choice theory* aims to investigate the rationality of collective decisions. In particular, in this field of study, the focus is often placed on which effects collective decision rules (e.g., simple majority vs. qualifies majority) or constitutional settings have on the rationality of social choice. In chapter three, we present main results of this branch of academic research.

In summary, the following figure shows main pillars of a theory of rationality.

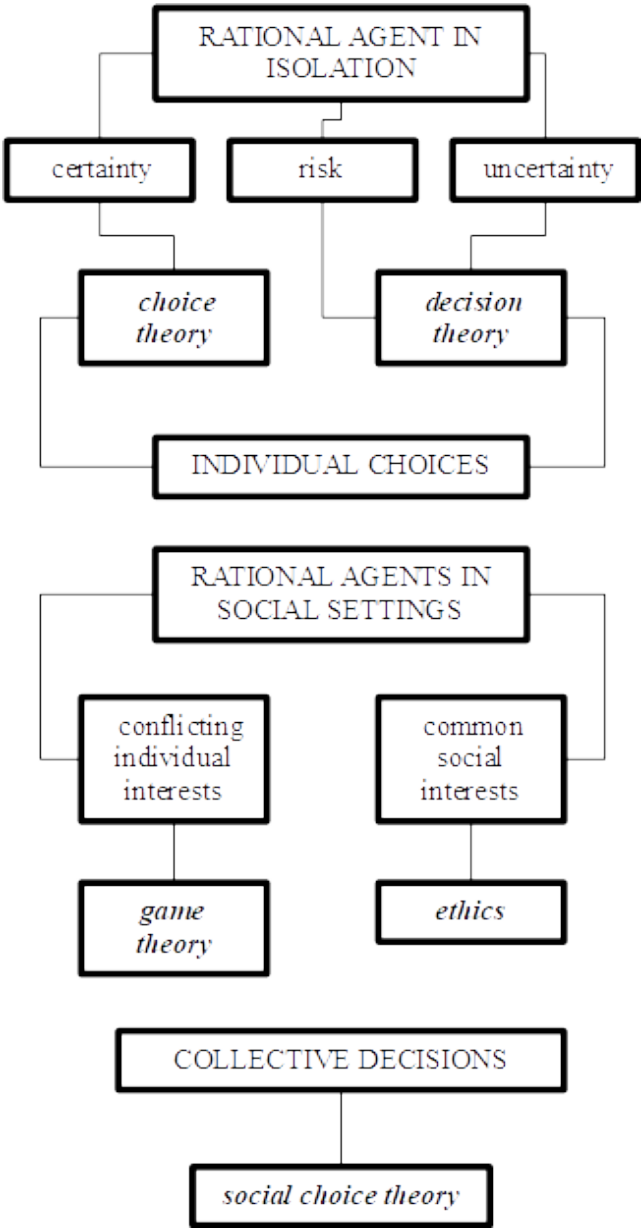

FIGURE 1: FROM INDIVIDUAL CHOICES TO COLLECTIVE DECISIONS

The Focus of the Book

We have now enough elements to specify the type of rationality on which the book focuses. As we have already pointed out, we restrict our presentation to individual and collective decisions in conditions of certainty. Moreover, we analyze intentional choices, or decisions, made using instrumental rationality. Instrumentally-rational behavior is examined, from both a substantive perspective and a procedural one, through the lens of methodological individualism. Even in the case of collective decisions, we consider rational individual agents who form committees or assemblies, but think, and vote, individualistically. Hence, according to the history of science, the approach of the book is *Humian-Weberian-Popperian* (HWP). Obviously, this is only one of possible analytical standpoints in the analysis of rational behavior. As recent debates have clearly emphasized, alternative notions of *ecological* rationality, *evolutionary* rationality or *we-rationality* can be very useful for addressing several important issues like environmental conservation or the management of common goods.[11] All these latest paradigms of rationality are out of the scope of this essay.

Instrumental Rationality and Utilitarianism

In models and theories of rational behavior, instrumental rationality often has been accompanied by the political philosophy of *utilitarianism*. In this section, we briefly discuss the reasons lying behind such common association.

Utilitarianism originated in United Kingdom during the XVIII century as a philosophical movement for political and moral reforms.[12] Utilitarian thinkers aimed to subject moral customs, social norms, political institutions, collective decisions, and so on, to the rule of *social utility*. Existing social practices, norms or institutions had to be reformed consistently with the maximization of social utility defined as the sum of *individual utilities*. At the beginning, individual utility was roughly conceived as the degree of satisfaction, happiness or pleasure enjoyed by an individual agent, but, very soon, the concept was filled in terms of *hedonic welfare*. The role played by marginalist economics was fundamental in this transition. Since the very beginning of the marginalist revolution in economics, the notions of ordinal and cardinal utility have been widely debated by economists.

[11] On this debate see Evan (2017).
[12] For a detailed analysis of utilitarianism see Kymlicka (1990).

The problem was whether it proved necessary to endow agents with a 'numerical', or measurable, utility function over their choice set (consumption set in demand theory).

For the first generation of marginalist economists, utility was a psycho-physical entity, and numerical hedonistic measures of pleasure/pain (*utils*) were assigned to different objects of choice. From this viewpoint, namely *cardinalism*, it was possible to have an objective measure of utility and comparable individual quantitative utilities. With the decline of the scientific project of measuring the intensity of sensations, dissatisfaction with this concept of quantitative, objectively measurable, utility grew.

Firstly, Pareto introduces a famous neologism, *ophelimity*, from the Greek word *ophelimós*, to distinguish between the objective utility of things and the subjective, psychological utility derived from needs or desires-satisfaction. Secondly, in his writings, Hicks spreads the idea that

> We have now to inquire whether a full theory of consumer's demand (…) cannot be built up from the assumption of a *scale of preference*. In constructing such a theory it will be necessary every time to reject any concept which is at all dependent on quantitative utility, so that it cannot be derived from the indifference map alone [...] we start off from the indifference map alone; nothing more can be allowed (Hicks (1939), p.21).

Hicks do not believe in cardinalism. He is persuaded that

> The ideal [agent] chooses that alternative, out of the various alternatives open to him, which he most prefers, or *ranks* more highly (Hicks (1956), p.58)

Therefore, ordinal utility is defined as the representation of a *preference ranking*, and the basic notion in choice models become that of individual *preference relation*, i.e., an order relation between alternatives that simply ranks options. Such a preference scale not only does not need to be cardinal, but as Samuelson (1947) argued, in order to have exclusively ordinal utility functions, preferences also must be exclusively ordinal (*ordinalism*). Following Mongin and d'Aspremont (1998), it is possible to summarize the core of ordinalism by three tenets: (*i*) utility functions are only a formal representation of preference orderings; (*ii*) preferences are merely a disposition to rank the possible options; (*iii*) choices can inform the observer on how preferences order options.

Hence, according to *ordinalist utilitarianism*, the satisfaction of preferences determines individual welfare/utility and the aggregation of utility levels

among agents quantifies the social utility, promptly renamed *social welfare*. The notion of instrumental rationality can be easily re-framed in terms of preference satisfaction. An individual agent is rational, in instrumental terms, if he/she selects options or/and actions which maximize preferences satisfaction, that is, which maximize ordinal utility. Moreover, a collective decision can be said rational only if it achieves the maximal level of social welfare. When the theory of rational behavior follows ordinalist utilitarianism, it is named *utility theory*.

Neo-Utilitarian Ethics

Utility theory also provides a quite simple rationality-oriented ethics.[13] According to the works of John Harsanyi, utilitarian ethics and instrumental rationality can be reciprocally connected through the principle of *equiprobability of moral judgments*. The intuition lying behind the principle is reasonable: in front of ethical problems, each individual has to reason *as if* he/she does not know his/her position in society, that is, considering an equal probability of occupying the social position of other agents. If the individual is rational, he/she will select the social outcome which grants the maximal expected utility, and the latter will correspond to the average utility level of all agents. Thus, Harsanyi's *neo-utilitarianism* suggests that a social decision rule, or a social state, is rationally justified, and ethically good, only if it entails the maximization of expected utility. As we will see in the third chapter, this ethical principle is deeply different from the one proposed by John Ralws' *A Theory of Justice*.

Instrumental Rationality and Hard Choices

There are situations in which the maximization principle of instrumental rationality is useless. Levi (1986) introduces the term *hard choices* for labeling choice problems in which values involved are both in conflict and difficult to weigh one with respect to others, i.e., to "resolve" their conflicting pulls is *hard*. This is the case of conflicts of justice, or moral dilemmas, which cannot be solved through the logic of maximization. In front of hard choices, individuals cannot maximize anything, and many incommensurable options exist. As Sen (2018) stresses, this incompleteness leaves unfinished many ethical reasoning about justice.

[13] See Harsanyi (1976).

Insofar as hard choices have been analyzed, contributions have been focused on how an agent should rationally deal with a hard choice. The reason for this is straightforward: even if from a theoretical perspective *indecisiveness* depicts well hard choice situations, in real-world circumstances it is not enough to invoke it, since a decision has to be taken, even if of radical, or horrendous, nature. Several possibilities have been suggested in the literature; let us briefly mention some of them.

A first possibility is to replace maximizing behavior with a procedure of satisfying choice. If this does not work for solving the dilemma, hence it is possible to introduce the *value-admissibility* of options, i.e., to distinguish between alternatives which are, and are not, forbidden according to the agent's values (*v-admissibility*).[14] Deleting no admissible options, and consistently reducing the decision problem, can make possible a rational choice. If also such attempt is fruitless, another possibility is: intersecting different value commitments to obtain an incomplete preference ranking of choice options and then applying to this last ranking the logic of maximization.[15] If also this last try does not make possible the choice, instrumental rationality must be dropped.

Instrumental Rationality and the Irrationality of Democracy

In the third chapter, we present several *impossibility results* of social choice theory. These theorems describe one clear difficulty/impossibility of systems of democratic decision: being rational. Democratic decisions based on voting procedures or collective decision rules are condemned to be incoherent, cyclic and inconsistent, all traits that reveal their irrationality. As we will show, technically speaking, the difficulty lies in the aggregation of different individual preferences in a clear and consistent social preference. Under very restrictive assumptions, such rational aggregation is feasible; when these conditions are weakened, the "irrationality of democracy" emerges.

The application of the idea of instrumental rationality to the functioning of democratic systems collides with an interesting scandal (from the Greek word *skàndalon*, i.e., obstacle). On the one hand, politicians are instrumentally-rational only if they maximize the number of votes they get in elections; on the other hand, citizens/voters are instrumentally-rational

[14] See Levi (1986).
[15] See Sen (2004).

only if they vote for the politician, or party, that allows them to maximize their preferences satisfaction. Nevertheless, when politicians form committees or parliaments in which they have to vote, or when citizens vote in elections, their individual rationality yield socially-irrational outcomes and policies. That is to say, by using an instrumental notion of rational behavior, the obstacle of "the irrationality of democracy" seems difficult to overcome.

The Philosophy of Rational Choice

There are three distinct philosophical assumptions behind a traditionally formulated rational choice (or utility) theory:[16]

(A1) rational behavior is regular and can be seen as maximizing behavior with a well-defined maximand;

(A2) the maximand is the self-interest (utility/preference satisfaction) of the agent;

(A3) the self-interest of the person is self-centered.

These are hence combined, by rationality theorists, with three commonly accepted claims about the individual "self":

(Γ1) *self-centered welfare*: An agent's welfare depends only on his/her own utility;

(Γ2) *self-welfare goal*: The agent's goal is to maximize his/her own welfare;

(Γ3) *self-goal choice*: An agent's choices are based entirely on the pursuit of the self-welfare goal.

By accepting the validity of both triples, (A1)-(A3) and (Γ1)-(Γ3), without defeaters, human rational behavior can be viewed as involving agents who

> Maximize their utility from a stable set of preferences and accumulate an optimal amount of information and other inputs in a variety of markets (Becker (1976), p.14).

A large part of contemporary economics, inspired by the pioneering contribution of Paul Samuelson, the theory of "revealed preferences", is

[16] See Sen (2002), pp.30-4.

based on the above description of rational behavior. [17]

By questioning the truthfulness of above assumptions and claims, and dropping some of them, it is possible to build less narrowly formulated theories of rational behavior.

First, by rejecting (A1), we get theories of *bounded rationality* like that originally suggested by Herbert Simon.[18] In this case, various elements (imperfect information, opacity of decision problems, imperfect computational abilities etc.) do not make it possible for individuals to be full maximizers. The evidence that individuals depart from maximization and look for satisfying choices has been delivered by many empirical works.

Second, by rejecting (A3), we have theories of non-self-centered rationality like the one suggested by Jolls *et al.* (1998). In these models, the central idea is that of *bounded self-interest*, i.e., the possibility of others-regarding concerns, concerns about the fairness of processes and the like. With bounded self-interest,

> People care about being treated fairly and want to treat others fairly if those others are themselves behaving fairly (Jolls *et. al*. (1998), p.1479).

As Becker (1996) has clearly shown, utility theory, based on tastes and preferences, can easily incorporate the possibility of altruism, or non-self-centered concerns. Moreover, this extension to a notion of non-self-centered self-interest is perfectly consistent with the logic of *indirect utilitarianism*, i.e., not acting in utilitarian terms maximizes utility.[19] Hence, if (A3) is dropped (Γ1) cannot be maintained.

Third, by rejecting (A2), we need to drop all claims (Γ1)-(Γ3). In this case, we need to address the case of non-self-oriented goals and choices. For example, take the case of commitments and self-imposed constraints. A *commitment* is a reason for action, or choice, irrespective of any personal loss suffered, even that deriving from the failure of the person's commitment.[20] From this perspective, a rational agent can select an action/option not because it makes him/her better off, but because he/she is committed to a value which reasonably asks for that action. There is room, in this case, for rational agents with goals other than their own welfare and

[17] See Samuelson (1938).
[18] The pioneering contribution on bounded rationality is Simon (1955).
[19] See Brink (1986).
[20] See Sen (2002) p.35.

with moral values other than the maximum fulfilment of their goals. In those cases, only a reasoned scrutiny and assessment of individual values, and social contexts of choice, will reveal what is being maximized. Hence, the diversity of possible "*reasons of choice*"[21] entails that narrowly formulated

> Maximizing behavior can sometimes be patently stupid and lacking in reasoned assessment […]. Rationality cannot be just an instrumental requirement for the pursuit of some given – and unscrutinized – set of objects and values (Sen (2002), p.39).

Accordingly, reasons of choice, not only *the fact* of choice make a difference. On the last point, Amartya Sen provides an apologue that we recall below.[22]

Example: A quarrel between two children concerning two apples is going on. Child 1 asks child 2 to take one of the two apples, and 2 immediately takes the bigger of the two. Child 1 grumbles: "You are so grabby. Given the first pick, I would have chosen the smaller apple". To this, child 2 replies: "So what are you grumbling about, you have got the one you would have chosen!"

In chapter two, we present several tools and categories which are useful in the building of paradigms of *weak rationality* open to commitment, self-imposed constraints and not self-centered choices.

Finally, there is the case of *bounded willpower*. In this scenario, problems of weakness of will, insufficient self-command or loss of self-control determine a departure of actual behavior from the dictates of instrumental rationality.[23] Now agents have different psychological interpretations, or reactions, to the situation they face and a relevant difference between the rational behavior suggested by the *planner*-self and that chosen by *doer*-selves may emerge.[24] Obviously, in these models, frequently discussed issues are *preferences reversal* and the *dynamic inconsistency* of choice. In short, when a smaller\earlier reward and a delayed\larger one are distant in time, individuals rationally prefer the delayed\larger one, but, when rewards get closer in time, they reverse their preferences. Arguably, such a reversal can be explained through hyperbolic functions used by individuals

[21] See Sen (2002) p.50.
[22] See Sen (1977).
[23] The pioneering contribution on these issues is Schelling (1984).
[24] See on the issue Benabou and Pycia (2002).

to discount the future: when the smaller reward gets closer in time, agents judge it better than the larger\delayed one.[25] Alternatively, it is possible to explain dynamically-inconsistent preferences in terms of unstable equilibria between sub-personal selves identified with neural sites. These sites are conceived as agents that interact to bring about a given choice behavior, and the brain is characterized as equilibrium of the strategic interaction between sites.[26]

Differently from the latter approaches, that can be grouped under the label of models of *quasi-rationality*,[27]

> [Traditional] rational choice theory is consistent with nonpsychological interpretations that in some contexts are more plausible (Satz and Ferejohn (1994), p.71).

Rationality and Freedom

"A man has free choice to the extent that he is rational". The quote of Aquinas in exergue could be used to explain in a sentence the main tie between Karl Popper's philosophy of science and his political thought. In Popper's view, obscurantism, ideologies, false historicism, and the like, are all grounded by irrationality and able to reduce human beings to irrational slaves. According to Amartya Sen, 'there is a *reciprocal relationship* between rationality and freedom': on the one hand, we cannot deny that 'there is a basic use of rational assessment in appraising freedom'; on the other hand, rationality depends on freedom 'because without some freedom of choice, the idea of rational choice would be quite vacuous.' [28]

As we will see in chapter three, in social choice theory the idea of liberty has been mainly developed in terms of *opportunity*, i.e., what an agent can choose to achieve in his/her private life. Freedom is conceived as the opportunity of not choosing x if y is preferred to it (and available); a power of the individual that is one of the basic principles of *political liberalism*.[29]

---

[25] This is the route followed by the so-called *pico-economic* approach. See Ainslie (1992).
[26] This is the route followed by the so-called *neuro-economics*. See Ross *et al.* (2007).
[27] The term was introduced by Richard Thaler. See Thaler (1991).
[28] See Sen (2002) p.5.
[29] On main features of political liberalism see Ralws (1995).

The individual freedom to prefer a particular private life has been named, in this strand of literature, *minimal liberty*. About that, several results, which we review in the third part of the book, have addressed to so-called "*Liberal Paradox*".[30] The paradox lies in the impossibility of getting unanimous social decisions which fulfil a minimal liberty requirement. Thus, unanimously accepted, or constitutionally accepted, liberal rights end to make more difficult, not less difficult, collective decisions. One possible reaction to the Liberal Paradox is the one which tries to refine rights systems, constitutions and collective decision rules in order to achieve effective social choices. In contrast, another reaction is seeing the paradox as the inevitable consequence of using a preference-driven notion of instrumental rationality.

Morality and Rational Preferences

One crucial problem of a preference-based utility theory, and in general of utilitarianism, is that of morally-unjustified, unreasonable, and hence morally-illegitimate, individual preferences. For example, if the recognition of right α (say, being not heterosexual) gives large utility to a small number of individuals and a larger disutility to the majority of people, is, or not, the preference for not recognizing α of majority's members legitimate? Or, put it differently, what if in order to maximize social utility, it is necessary to violate equality of treatment and fairness, to negate some people's rights?

To answer to these questions, the distinction between *personal* preferences and *external* preferences, originally suggested by Dworkin (1977), is useful. Personal preferences are those appraising goods, actions or opportunities that the individual wishes to dispose of. By contrast, external preferences are those concerning goods, rights and actions that an individual wishes others to have. Sometimes, external preferences are the result of prejudice and discrimination, in other cases they question what is legitimate for some groups of people.

In principle, utilitarianism does not accept the view that some preferences are illegitimate: there can be no criteria which precede the maximization of social welfare and a person's legitimate rights, or opportunities, are established by such a calculation. By proceeding in this way, however, utilitarian theories end up denying one of their basic principles: *universalism*, i.e., all individuals have moral relevance and, for this reason, all individual

[30] See the symposium on the paradox in *Analyze & Kritik*, September 1996.