UNIVERSITÀ DEGLI STUDI DI TRIESTE
Sede Amministrativa del Dottorato di Ricerca

XXI CICLO DEL
DOTTORATO DI RICERCA IN
FILOSOFIA

STANDARDS OF RATIONALITY AND CONTEXT
(Settore scientifico-disciplinare: M-FIL/05 FILOSOFIA E TEORIA DEI LINGUAGGI)

DOTTORANDO
PAOLO LABINAZ

COORDINATRICE DEL COLLEGIO DEI DOCENTI
CHIAR.MA PROF.SSA
MARINA SBISÀ
UNIVERSITÀ DI TRIESTE

TUTORE
CHIAR.MA PROF.SSA
MARINA SBISÀ
UNIVERSITÀ DI TRIESTE

RELATRICE
CHIAR.MA PROF.SSA
MARINA SBISÀ
UNIVERSITÀ DI TRIESTE

ANNO ACCADEMICO 2008-2009
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter 1</strong> <em>Questioning the standard picture of rationality</em></td>
<td></td>
</tr>
<tr>
<td><em>Introduction</em></td>
<td>5</td>
</tr>
<tr>
<td>1. <em>The background of the rationality question</em></td>
<td>6</td>
</tr>
<tr>
<td>1.1 Reasoning and rationality: a preliminary distinction</td>
<td>6</td>
</tr>
<tr>
<td>1.2 Two general conceptions of rationality</td>
<td>9</td>
</tr>
<tr>
<td>2. <em>The deontological account of rationality</em></td>
<td>12</td>
</tr>
<tr>
<td>2.1 The Standard Picture of Rationality</td>
<td>13</td>
</tr>
<tr>
<td>2.1.1 Deductive logic</td>
<td></td>
</tr>
<tr>
<td>2.1.2 Standard Probability Theory: the Bayesian (or subjective) interpretation</td>
<td></td>
</tr>
<tr>
<td>2.1.3 Expected Utility Theory as individual decision theory</td>
<td></td>
</tr>
<tr>
<td>2.2 The virtues of the standard picture of rationality</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Deontologism and normativity</td>
<td>18</td>
</tr>
<tr>
<td>2.4 The rationality condition for individuals</td>
<td>20</td>
</tr>
<tr>
<td>3. <em>The rationality debate</em></td>
<td>22</td>
</tr>
<tr>
<td>3.1 The psychology of reasoning: three experimental cases</td>
<td>22</td>
</tr>
<tr>
<td>3.1.1 The Wason Selection Task</td>
<td></td>
</tr>
<tr>
<td>3.1.2 The Linda Problem</td>
<td></td>
</tr>
<tr>
<td>3.1.3 The Asian Disease Problem</td>
<td></td>
</tr>
<tr>
<td>3.1.4 Simplicity vs. complexity</td>
<td></td>
</tr>
<tr>
<td>3.2 Rationality vs. Irrationality Theses</td>
<td>29</td>
</tr>
<tr>
<td>3.3 Systematic violations or occasional mistakes?</td>
<td>33</td>
</tr>
<tr>
<td>3.4 Heuristics and biases</td>
<td>34</td>
</tr>
<tr>
<td>3.5 Two general arguments in defence of human rationality</td>
<td>37</td>
</tr>
<tr>
<td>3.5.1 The necessary presumption of rationality</td>
<td></td>
</tr>
<tr>
<td>3.5.2 The alternative task interpretation argument</td>
<td></td>
</tr>
<tr>
<td>4. <em>The Standard Picture of Rationality in question</em></td>
<td>44</td>
</tr>
<tr>
<td>4.1 Which normative model is the right one?</td>
<td>44</td>
</tr>
<tr>
<td>4.2 Universality and necessity in rationality assessments</td>
<td>46</td>
</tr>
</tbody>
</table>
Chapter 2  
Non-standard deontological approaches to rationality assessment

Introduction

1. Rationality and ordinary intuitions
   1.1 Ordinary intuitions and reflective equilibrium
   1.2 Why do ordinary intuitions matter?
   1.3 Ordinary intuitions and probabilistic reasoning
   1.4 Ordinary intuitions and relativism
   1.5 Wide reflective equilibrium and experts’ intuitions

2. Minimal Rationality
   2.1 The human finitary predicament
   2.2 To be minimally rational
   2.3 Relevance and context

3. Non-standard deontological conceptions and rationality assessment
   3.1 The rationality condition for individuals
   3.2 What goes wrong with intuition-based and minimal rationality?
   3.3 From normative to prescriptive models of rationality
   3.4 The blind spot of any deontological account

4. Concluding remarks

Chapter 3  
Consequentialism and adaptation

Introduction

1. Consequentialism: from the standard picture of rationality to bounded rationality
   1.1 Reasoning and goals
   1.2 Consequentialism and the standard picture of rationality
   1.3 Bounded rationality
      1.3.1 Maximizing or satisficing?
      1.3.2 Is bounded rationality really normative?
   1.4 From maximization to satisficing: a final balance

2. Rationality and adaptation
   2.1 Adaptations as truth-conducive
   2.2 Evolutionary psychology
   2.3 Reasoning, rationality and adaptation

3. Ecological Rationality
   3.1 Ecological validity
3.2 The adaptive toolbox 106
3.3 Ecological rationality and normativity 108

4. The implications of adaptive and ecological rationality for the evaluative question 110
4.1 Adaptive rationality and the context of evaluation 110
4.2 Ecological rationality, environments and goals 112

5. Dual Rationality Theories 114
5.1 The two cognitive systems 115
5.2 Two kinds of rationality 117
5.3 The implications of dual rationality for the evaluative question 119

6. Concluding remarks 120

Chapter 4 Consequentialism and pragmatism

Introduction 123

1. From the standard analytic approach to epistemology to epistemic consequentialism 124
1.1 Epistemic consequentialism: an alternative way to approach epistemology 124
1.2 Epistemic goals: truth and beyond 126

2. Stich’s Epistemic Pragmatism 128
2.1 Descriptive and normative cognitive pluralism 128
2.2 Normative cognitive pluralism: beyond truth as the main epistemic goal 135
2.3 Pragmatism and relativism 139
2.4 Epistemic Pragmatism’s implications for the evaluative question 141
2.5 Against Epistemic Pragmatism 143
   2.5.1 The value of truth
   2.5.2 Pragmatism and value-based assessments

3. Strategic Reliabilism 147
3.1 Strategic Reliabilism as a naturalistic approach to epistemology 147
3.2 Evidence-based criteria of evaluation 148
3.3 Linking Ameliorative Psychology and Strategic Reliabilism: the Aristotelian Principle 150
3.4 Truth and good outcomes 151
3.5 Against Strategic Reliabilism 153
   3.5.1 The Aristotelian Principle
   3.5.2 Well-being and good outcomes

4. Cognitive and situational context 156
4.1 Pragmatist approaches to rationality assessment 156
Chapter 5  

Towards a situated consequentialism

Introduction 164

1. Situational context and objectivity 165
   1.1 Why does context matter in reasoning? 166
   1.2 Context and conversation 167
   1.3 Towards an objective notion of context 171
   1.4 Rationality assessment and objective context 174

2. A context-sensitive consequentialist approach
   to rationality assessment 176
   2.1 Goals, relevant information and reasoning strategies 177
   2.2 A two-step normative framework for rationality assessment 180
   2.3 Situatedness and normativity 184

3. Applying the context-sensitive consequentialist approach:
   the case of the Wason Selection Task 187
   3.1 Standard version of the selection task and hypothesis testing 188
   3.2 A brief history of the studies on the selection task 189
   3.3 Focusing on the indicative version: comprehension and probability 192
      3.3.1 Relevance-based explanation
      3.3.2 Probability and Optimal Data Selection
   3.4 A two-step evaluation of subjects’ reasoning performances 199

4. Concluding remarks 205

Chapter 6  

Concluding Remarks 207

References 215
This dissertation deals with the debate among analytic philosophers and cognitive psychologists concerning the implications that experimental findings on human reasoning have for the evaluation of human rationality, known as the rationality debate. My main claim is that many disputes in this debate could be made to disappear by relying on an appropriate notion of context.

Since the early seventies, several experimental researches on human reasoning have produced evidence that human beings diverge significantly from the classical normative standards of rationality, failing to solve apparently simple reasoning tasks. A debate has then arisen among analytic philosophers and cognitive psychologists about whether or not these experimental findings show that humans are basically irrational. Much of this controversy depends on different assumptions about which normative standards of rationality are taken to be the appropriate ones and what precisely is being assessed against them.

Throughout this work, I rely on a distinction between two different conceptions of rationality which correspond to two different, competing targets in rationality assessments: the deontological and the consequentialist conception. While the former states that to be rational means to reason systematically in accordance with a definite set of normative principles, dictated by an accepted normative model of rationality, the latter claims that to be rational means to reason in such a way so as to be likely to achieve effectively a goal or a
range of goals. On the basis of this distinction, I characterize four different approaches to rationality assessment, which I examine in the first four chapters: the standard and non-standard deontological approaches, the evolutionary consequentialist approach and the pragmatist consequentialist approach.

In the first chapter, I focus on the standard deontological approach to rationality assessment, according to which normative principles of rationality stem from deductive logic, standard probability theory and expected utility theory. As empirical data on human reasoning seem to show, however, subjects’ performances on reasoning tasks fall short of these normative principles. That may be either because people are indeed irrational or because the standard deontological approach is flawed and insufficient for assessing human reasoning. I opt for the second option and argue that what is at stake is not the validity of normative principles in the abstract but, rather, the way they are applied to concrete situations.

In the second chapter, I turn to two non-standard deontological approaches to rationality assessment, developed by Jonathan L. Cohen and Christopher Cherniak respectively. Disagreeing with the standard approach, Cherniak and Cohen give two distinct characterizations of the normative principles against which to assess human reasoning. According to Cohen, the normative standards should ultimately be based on people’s ordinary intuitions rather than on formal theories, whereas Cherniak, holding that it is impossible for humans to perform all the inferences the standard deontological approach requires, redefines what it means to be rational on the basis of the limitations of the human mind. After proceeding to a critical assessment of these deontological accounts, I hold that both are inadequate for assessing human reasoning. In the final part of the chapter, I argue that any deontological account, either standard or non-standard, is flawed as the idea that complying with a given set of normative principles is what characterizes good reasoning does not provide an explanation of why people should reason well.
In the third and the fourth chapter, I examine two consequentialist approaches to rationality assessment, grounded respectively on evolutionary and pragmatist considerations. In the third chapter, I examine how evolutionary psychology is used to support the consequentialist conception of rationality. According to the evolutionary and ecological views of rationality, evolution, through natural selection, was directed towards increasing reproductive success and every part of the human body, including the brain, has been selected because it is a good instrument for achieving that goal. Consequently, human reasoning is seen as having evolved to the promotion, direct or indirect, of reproductive success. However, this brand of consequentialism falls short of providing an adequate approach to rationality assessment. In particular, consequentialist accounts based on evolutionary considerations provide a very limited framework for assessing human rationality as the kinds of subject’s goal they take into account are always linked, directly or indirectly, to inclusive fitness.

In the fourth chapter, as an alternative to the evolutionary approach, I examine two consequentialist approaches to rationality assessment inspired in different ways by pragmatism, i.e., Stephen Stich’s Epistemic Pragmatism and Michael Bishop and John Trout’s Strategic Reliabilism. I argue that, by focusing exclusively on matters such as the cognitive resources available to reasoners and the value or significance they attach to the goal they pursue, pragmatists tend to neglect the distinction between the situational constraints and the cognitive constraints under which reasoners operate and consequently fail to distinguish their distinct roles in rationality assessment. Finally, I point out at the vacuity of any consequentialist approach not taking into serious consideration the situational constraints in human reasoning.

In the fifth chapter, after considering a further approach to the study of human reasoning, which takes into account its conversational dimension and its relationship to context, I propose a way of approaching rationality assessment which I call the context-sensitive
consequentialist approach. Starting from some considerations as regards the situatedness of speech acts, inspired by the work of J.L. Austin, I oppose an objective notion of context to the cognitive one assumed by proponents of the conversational approach to rationality assessment, and then explore the implications that the adoption of an objective notion of context would have in this field. By focusing on the strong interdependence between the particular structure of a reasoning task and its understanding on the part of the subjects, I propose a two-step normative framework for situationally establishing the legitimacy of the subjects’ interpretations of the task and the normative appropriateness of their responses and apply it to the data on human reasoning obtained from the standard version of Wason’s selection task. Finally, I surmise that the two-step normative framework of the context-sensitive consequentialist approach has a potential for application to other types of reasoning tasks and for use in combination with other approaches, such as evolutionary and pragmatist ones, towards an overarching framework for the assessment of human rationality.

**Acknowledgements.** I would like to thank Prof. Marina Sbisà for her constant supervision and invaluable support throughout my PhD studies. My gratitude goes as well to Prof. Stephen Laurence for supervising my research work during my stay at the University of Sheffield in fall 2008 and to Prof. Marcello Frixione for giving me many valuable suggestions to improve the dissertation. I am also grateful to all the members and PhD students of the Department of Philosophy, Languages and Literature of the University of Trieste.
Chapter 1 – Questioning the standard picture of rationality

Introduction

In this chapter, I begin by providing an overview of both theoretical issues and empirical findings with regard to human reasoning and rationality. I explore first some controversial issues regarding the debate on the role and status of normative assumptions about rationality in philosophical theorizing and then the implications that a set of well-known experimental results on human reasoning has had on this debate. These experimental results show that subjects’ performances on apparently simple reasoning tasks frequently depart from the classical normative principles of rationality, yielding normatively inappropriate responses. A lively debate has taken place among analytic philosophers and cognitive psychologists about whether or not these experimental findings show that humans are basically irrational. The researchers involved in this debate divide into two groups on the basis of their claims regarding the extent to which the principles being in human reasoning competence correspond to those belonging to the so-called standard picture of rationality.

On the one hand, some psychologists and philosophers have tried to elaborate upon the empirical findings on human reasoning and to conclude that human beings are basically irrational. On the other hand, other researchers also assuming that the standard picture of rationality holds, have argued that most instances of apparently irrational behaviour can easily be explained away by appealing to various (cognitive and non-cognitive) factors: human beings should be considered as rational, after all. In my view, however, these claims about human rationality depend on the researchers’ reliance on a questionable normative
framework for assessing human reasoning, which I call into question at the end of the chapter.

1. The background of the rationality question

In this section, I preliminarily analyze the general background against which the rationality question can be examined. I introduce three questions with regard to the relation between human reasoning and rationality, which I call the descriptive, the normative and the evaluative question (following Samuels, Stich & Faucher 2004), and stress the fundamental role played by the evaluative question in the debate about the nature of human rationality. Then, I pass on to introduce some widely accepted distinctions among conceptions of rationality. Among these distinctions, I focus on a particular one (proposed by Samuels, Stich & Faucher 2004) that identifies two different conceptions of rationality, the deontological and the consequentialist conception, which correspond to two different, competing targets of rationality assessments.

1.1 Reasoning and rationality: a preliminary distinction

Since the early seventies, a lively debate, which is known as the rationality debate, has arisen among cognitive psychologists and analytic philosophers as a consequence of the great deal of experimental findings which seem to reveal systematic errors in human reasoning (see, e.g., Evans et. al. 1993; Gilovich et al. 2002; Kahneman et al. 1982). Inasmuch as rationality has been identified with the capacity to reason well and effectively,¹ these systematic failures have prompted many researchers to investigate the

¹ Good reasoning is usually considered as the central example of a rational process. It is widely agreed, however, that other cognitive activities, such as perception and memory, may lead to rational beliefs and decisions, in ways that do not necessarily require reasoning (see, e.g., Audi 2004). My focus here will be only on reasoning.
extent to which humans can be considered as rational. If we assume, as Aristotle and most other Western philosophers did,\(^2\) that possessing rationality is the feature that distinguishes humans from other animal species, the experimental findings on human reasoning seem to call into question the traditional characterization of man as the rational animal: people do not usually behave as rationally as philosophers might expect in a wide variety of situations calling for deductive and inductive reasoning, probabilistic judgment, or decision making.\(^3\) What can we conclude from these empirical data? Are humans really irrational? But also, what kind of issues are we faced with when we pose such questions? An overall reply to the above mentioned questions requires some preliminary clarifications. Before all, we need to start by examining how the relationship between human reasoning and rationality has been and is now conceived in cognitive psychology and analytic philosophy, particularly epistemology. For many decades, it has been assumed that (i) cognitive psychologists investigating human reasoning do not need to take into account normative issues about rationality and (ii) philosophers may investigate rationality without examining the results of those empirical studies (see, e.g., Cohen 1981; Evans 1982). This received view of the relationship between reasoning and rationality holds that: (a) human reasoning conceived as an activity performed by the human mind is the subject matter of cognitive psychology; (b) the study of rationality requires a normative analysis of reasoning and is thus the subject matter of a philosophical analysis; (c) the empirical studies on human reasoning and the normative approach to rationality have different aims and methods and so must therefore be conducted independently within the disciplines of cognitive psychology and philosophy of

---

\(^2\) Aristotle’s standpoint about this topic is to be found in his *De Anima* (*On the Soul*). According to him, what distinguishes humans from other animal species is the fact that only the former possess a capacity to think and reason, i.e. they possess a rational soul.

\(^3\) By “human reasoning” I mean a more or less conscious cognitive activity that involves drawing conclusions on the basis of some evidence, assumptions or reasons. Human reasoning can be directed at forming beliefs, arriving at a judgement, making a decision, solving a problem and the likes. Characterized in such a way, it includes deductive and inductive reasoning, probabilistic judgment, and decision making. Reasoning has to be distinguished from arguments which normally are characterized as sets of statements “that one person offers to another in the attempt to induce that other person to accept some conclusions” (Pinto 1995: 32).
logic or epistemology respectively (see Moshman 1987). As a consequence, it was held to be completely irrelevant to investigate whether and to what extent the actual reasoning of ordinary people conforms to what is dictated by the normative models of rationality. More recently, many epistemologists have shown considerable interest in the empirical study of human reasoning and have begun to consider psychological findings in their own theorizing (see, e.g., Goldman 1986; Kornblith 1993; Stich 1990). At the same time, most psychological studies on human reasoning directly take into account the relation of subjects’ reasoning performances to normative standards of rationality. Nowadays, there are three general and interrelated questions which need to be taken into account when the relationship between human reasoning and rationality is examined: the descriptive, the normative (and meta-normative) and the evaluative question (Samuels et al. 2004: 131).

The first question calls for a psychological account of how humans reason, which usually includes a theory of (i) their reasoning competence and (ii) their reasoning performances. The normative question calls for a general normative model of how rational subjects should or ought to reason, which includes (i) a set of normative principles of reasoning (following which makes human thought rational) and (ii) a meta-normative account of how these principles can be justified and (if necessary) revised (Hardy-Vallée & Thagard 2008: 174). Finally, the evaluative question calls for a general evaluative framework for assessing human reasoning. It concerns the different ways of determining the extent to which a particular instance of reasoning accords with a set of appropriate normative standards.

Insofar as you are interested in the question of human rationality, you should take into consideration all these three questions as they are inextricably intertwined: one starts with a collection of data about performances on a reasoning task and an appropriate normative standard for evaluation, determine whether those performances comply with the standard and, ultimately, thereby whether humans can be considered as rational. That is why any
attempt to approach human rationality should provide, at least implicitly, a partial response to all of these three questions. Looking at the literature on human reasoning and rationality, you can find several explicit attempts to approach the descriptive and the normative question. But, there are only few attempts to explore and discuss systematically how a general evaluative framework for assessing human reasoning should be characterized. As we will see, if two researchers assume the same normative standards for rationality but are committed to different evaluative frameworks for assessing human reasoning, that may lead them to argue for very different and competing claims about human rationality. So, the evaluative question plays a fundamental role in determining the bearing that this collection of experimental findings on human reasoning has on the rationality question. But in what terms are rationality assessments usually framed? What is the object of evaluation? What are the targets of the assessment? These questions require the introduction of some distinctions among conceptions of rationality. Indeed, as we will see, the researchers’ ways of assessing reasoning performances are strongly determined by the conception of rationality to which they are committed.

1.2 Two general conceptions of rationality

In the last decades, questions about human rationality have received an increasing amount of attention from social and scientific disciplines studying human behaviour. There remain, however, a lot of open questions about the conceptualization of rationality and its use: what kind of things are considered to constitute a rational phenomenon? When do questions about rationality arise? These are not simple and clear issues. Philosophers, psychologists and social scientists regularly use the term “rational” with reference to a wide variety of items. For example, John Elster (1983: 1) has proposed a very long list which includes “beliefs, preferences, choices or decisions, actions, behavioural patterns, persons, even
collectivities and institutions”. When dealing with human behaviour, scholars belonging to
different disciplines like philosophy or psychology use the term “rationality” in slightly
different ways; some focus on cognitive processes and others on behavioral dispositions in
accordance with their different assumptions and aims. So, there is the danger that in
interdisciplinary discussions on rationality several misunderstandings may arise. To avoid
this danger, some further distinctions may be helpful in approaching the rationality
question. In particular, I would like to delimit what precisely is being referred to with the
term “rational” in my work. I am here concerned with neither institutional rationality nor
collective rationality (see, e.g., respectively, Smokler 1983; McMahon 2001). Rather, my
interest here is on what makes belief, judgment, and decision rational. With regard to these
items, classically philosophers have divided rationality into what may be called epistemic
or theoretical rationality and practical rationality: “whereas theoretical or epistemic
rationality is concerned with what it is rational to believe, and sometimes with rational
degrees of belief, practical rationality is concerned with what it is rational to do, or intend
or desire to do” (Mele & Rawling 2004b: 3). Both theoretical and practical rationality are
regarded as demanding reasoning: while theoretical reasoning is aimed at producing true
beliefs about the world, in practical reasoning people aim at selecting the best action to take
in pursuit of their goals. These two kinds of rationality are widely interconnected and even
display some similarities. In particular, both epistemic and practical rationality can be
distinguished into two distinct forms, which I call coherence-rationality and process-
rationality (for a similar distinction, see Kahneman 2000). With regard to theoretical
rationality, while the first form refers to the consistency and coherence of a belief-set, the
second one is applied to the processes of belief formation and revision. In the same vein,
when characterized in terms of coherence, practical rationality refers to the formal
consistency among one’s preferences, considered as a necessary condition for making
decisions that are rational. With regard to the second form, practical rationality is concerned
with deciding what to do to achieve one’s goal, that is, figuring out the appropriate means
to attain that goal. So, coherence-rationality is concerned with the internal coherence of the
subject’s systems of preferences and of beliefs, regardless of her cognitive processes and
goals. According to Donald Davidson (1982: 290), for example, the fundamental
philosophical question about rationality concerns the failure of consistency among the
beliefs or preferences of a single individual rather than the failure of someone to do or
believe what is taken to be rational by others. However, the researchers involved in the
rationality debate are mainly concerned with process-rationality both in its theoretical and
practical form. Their objects of evaluation are instances of reasoning which are aimed at
inferring conclusions and making judgments or decisions. Even if these researchers are
referring to the same “object”, that is, an instance of reasoning, we can distinguish the
targets of their assessments according to two general conceptions of rationality: the
deontological and the consequentialist one (Samuels et al. 2004: 165-166). According to
the former, to be rational means to reason systematically in accordance with a definite set
of normative principles, dictated by an accepted normative model of rationality, whereas,
according to the latter, to be rational means to reason in such a way so as to be likely to
attain efficiently a goal or a range of goals. Both these conceptions are assumed by very
different and competing accounts of rationality, such as those based on ordinary intuitions
and, respectively, evolutionary considerations. One the one hand, while aiming at two very
different targets, these conceptions agree on the fact that rationality requires normativity in
the sense that there must be a difference between doing a rational thing and doing an
irrational one. Indeed, to call something rational is not to describe it but rather to make an
assessment. As is argued by Bermúdez (2003: 245), such a characterization of rationality

4 Normativity is a central issue for many philosophical disciplines, particularly epistemology and the
philosophy of language no less than ethics and aesthetics.
presupposes the existence of a space of alternatives in which the right thing to do always contrasts with some other actions that could have been performed: it cannot be the case that anything people might do would qualify as rational. On the other hand, however, the deontological and consequentialist conceptions diverge on what is that makes the difference between good and bad reasoning. This fundamental difference raises some questions: what are the criteria we should use to distinguish what is rational from what is not? And to which domain of rationality do they belong? What gives them their normative force?

There is clear disagreement between researchers involved in the rationality debate as to whether rationality should be defined in terms of conformity to normative principles or in terms of achievement of goals. Though few of them overtly state which one these characterizations they favour, most researchers involved in the rationality debate have tacitly adopted a deontological approach to rationality assessment which is based on what Edward Stein (1996: 4) calls the *standard picture of rationality*. In the rest of the chapter, I will examine such a deontological approach to rationality assessment and the implications that its adoption has for what we have called the evaluative question.

2. The *standard deontological approach to rationality assessment*

In this section, I introduce what I call the standard deontological approach to rationality assessment. I first provide a general characterization of the *standard picture of rationality* on which such an approach is based and then examine what its normative principles are and where they come from. Finally, I distinguish between the normative standards of rationality against which to assess human reasoning, and what I will call the general requirement for rationality, which articulates the rationality condition for an individual.
2.1 The standard picture of rationality

Traditionally, it has been assumed that rationality demands stable and firm standards so that we can distinguish it from subjective and changeable opinions about what is the best thing to believe or do. If no such standards are available, the rationality question is likely to remain unresolved. For a long time, most analytic philosophers, cognitive psychologists and also economists have relied on what Edward Stein (1996: 4) calls the *standard picture of rationality* (*SPR*), according to which,

> to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth. If the *standard picture* of reasoning is right, principles of reasoning that are based on such rules are *normative principles of reasoning*, namely they are the principles we *ought* to reason in accordance with.

According to the supporters of this picture, to be rational means to reason well and effectively and if we ask educated people how we can reason well and effectively, the answer will be obvious, that is, it means to reason in accordance with the normative principles of reasoning that are derived from rules and axioms of deductive logic, standard probability theory and expected utility theory. Insofar as most researchers involved in the rationality debate have assumed that these normative standards are stable and firm, we might wonder why. In other words, we have to examine where the source of their normative force comes from and why a subject ought to follow them in order to count as rational.

As I understand the claims of its supporters, the *standard picture* is regarded as providing a set of *a priori* normative principles of rationality. If it is so, the source of its normative force can be explained by referring to what Renée Elio calls the Frege-Husserl view of logic (Elio 2002b: 5; see also Frege 1894; Husserl 1900). According to this view,
logical facts (theorems, axioms and principles of logic as well as, in our case, those of probability theory and expected utility theory) are considered to be true independent of any cognitive activities, not being determined by how the human mind works. Their normative force is taken to be universal and not contingent upon factual considerations. As a consequence, these normative principles are considered true regardless of whether humans believe them or not. That implies that the normative and the descriptive question are totally independent, as whatever we might discover about human reasoning will not give use any advice about how people should behave in order to count as rational. When human reasoning falls short of a given normative principle, there is always something wrong with the former, not with the latter. On this basis, the experimental results on human reasoning are relevant only in the descriptive and the evaluative question characterized above: while, in the first case, researchers use this evidence to develop a theory of how humans reason, in the second one, reasoning performances are assessed against *a priori* normative standards.

To sum up, the *standard picture* combines a deontological conception of rationality with a set of *a priori* normative standards. According to it, to reason rationally is to reason (explicitly or implicitly) in accordance with that set of *a priori* normative principles, and when we fail systematically to conform to them in reasoning, we can be charged with irrationality. However, as Stein’s characterization of the *standard picture* suggests, axioms or principles of the formal theories of reasoning are not themselves normative. The question is, therefore, how the norms of rationality may be derived from deductive logic, standard theory of probability and expected utility theory, and how they should be employed so as to assess human reasoning. That requires a further step. In order to explain that, I sketch out first the main characteristics of these formal theories of reasoning and then hypothesize how the supporters of the *standard picture* may derive the normative principles from them.

---

5 Supporters of the *standard picture* hold that the validity of its principles depends on mathematical and logical demonstrations. Being considered to be derived from these *a priori* sources, such standards are attributed an universal power.
2.1.1 Deductive Logic

As is well known, “logic” is a term which can refer to slightly different related issues, and so we should be very careful about what different logicians mean by using it. When researchers involved in the rationality debate talk about logic, they usually refer to classical deductive logic.\(^6\) In its standard characterization, deductive logic is concerned with entailment relations among propositions: if all the premises of a deductive argument are true the conclusion will be necessarily true and consequently it is not the case its conclusion may be false when all the premises of this argument are true - all the arguments based on true premises that are deductively valid are truth preserving. It is noteworthy that, on this definition, validity depends only on the arguments’ formal structure and not on the matter or meaning of the premises. Deductive logic provides a general formal framework that distinguishes between those argument structures that are truth-preserving and those which are not. Tellingly, as is stressed by Jonathan Evans (2002), the rationality debate has developed under the influence of the “deduction paradigm”, according to which to reason rationally is to correctly reason deductively.

2.1.2 Standard Probability Theory: the Bayesian (or subjective) interpretation

The normative status of probability theory is very controversial. There are different traditions which differ on how probability should be understood. I restrict this brief introduction to a particular interpretation of probability which has been widely assumed among the researchers involved in the rationality debate. By means of this interpretation, the probability of an event corresponds to one’s subjective degree of belief that the event will occur or has occurred. This interpretation is well-known as the Bayesian (or subjective)

---

\(^6\) Classical deductive logic has two main features: (i) it assumes the bivalence principle, that is, the claim that every proposition is either “true” or “false” and so there are no truth-value gaps and (ii) it includes truth-functional connectives (such as negation and conjunction) and quantifiers (such as universal and existential quantification) as its only logical operators. For a recent introduction to propositional and predicate logic see, e.g., Restall 2006.
view of probability. According to this view, the judgment of the likelihood that an event will occur expresses the degree of belief of a subject S in the occurrence of that event. In order to assign normatively appropriate probabilities, the subject S must satisfy three conditions: (i) she must have a set of probabilistic beliefs that is coherent, namely she must assign probabilities to her beliefs in a way that assures that they comply with Kolmogorov axioms (the basic mathematical axioms of the standard probability calculus);\(^7\) (ii) her set of probabilistic beliefs must be complete in the sense that she must assign a subjective probability to each proposition; (iii) S’s beliefs must be revised in accordance with her conditional probabilities (see Grüne-Yanoff 2007: 536-537).\(^8\) Within this probabilistic framework it is assumed that the subject S will conform to what is dictated by Bayes’ Theorem, which has to be used to update the conditional probability of an event or a hypothesis when new relevant information are given.\(^9\)

2.1.3 Expected Utility Theory as individual decision theory

According to expected utility theory (EUT), given a range of alternative options, the optimal choice is the one that maximizes one’s expected utility, wherein the value of an option amounts to the sum of the utilities of its outcomes multiplied by the probability of its occurrence. As is pointed out by David Velleman (2000: 144), this theory provides both a mathematical way to formalize the relations among one’s value, belief, and preference and a set of prescriptions for maximizing her expected utility. It always begins with a subject S

\(^7\) These axioms can be summarised as: (i) a probability is expressed as a number in the range 0 to 1; (ii) the sum of the probabilities of all the elementary events in a given situation equals to 1; (iii) for any collection of mutually exclusive events, the probability of their union, that is, that any of them occur amounts to the sum of their separate probabilities.

\(^8\) Conditional probabilities are expressed with the following notation \(P(\mid )\), and \(P(A\mid B)\) specifies the probability that \(A\), given that \(B\) has occurred.

\(^9\) Stated in terms of hypothesis testing, the definition specifies how a hypothesis should be revised in the light of new relevant data. Its formula is the following: \(P(H\mid d) = P(d\mid H)P(H)/P(d)\), provided that \(P(d) > 0\). That is to say, the probability of a hypothesis given the data \(P(H\mid d)\) (called the “posterior probability”) amounts to the probability of the data given that hypothesis \(P(d\mid H)\) multiplied by the current (“prior”) probability of the hypothesis \(P(H)\) on the basis of the likelihood of the data \(P(d)\).
and a range of alternative options available to her. These options are deemed to be uncertain as they may lead to different possible outcomes on the basis of whether certain conditions occur. In particular, the standard version of expected utility theory assumes that a subject \( S \) is faced with two alternative options, that is, she has binary preferences among the two alternative options. Provided that these binary preferences meet a set of formal requirements (depending on which axiomatisations we refer to)\(^{10}\) two functions will be available to \( S \), which respectively assess the probability of each possible outcome and identify the utility to be derived from it, in such a way that \( S \) will then choose the option that provides the optimal combination of utility and probability (see, e.g., Velleman 2000: 144-145).\(^{11}\) In doing so, the subject’s decision will be described and also justified as maximizing her expected utility.

### 2.2 The virtues of the standard picture

Why should we prefer the standard picture of rationality to any other competing pictures of rationality? Supporters of the standard picture claim that it possesses three main virtues: (i) it provides stable, \( a \ priori \) standards against which to assess human reasoning so that rationality can be distinguished from one’s own subjective opinions about what is the best thing to believe or do; (ii) it fits well with our intuitions about deductive logic, standard probability theory and expected utility theory and their close relationship to issues about how people ought to reason and make decisions; (iii) it captures a received idea about rationality, that is, its normative principles are universal and apply in any context. Deductive logic is perhaps the best-known example of universality. Depending only on its

---

\(^{10}\) The requirements which \( S \)’s preferences must satisfy are taken to be the axioms of the theory. Different theories assume different axioms, such as transitivity, invariance, continuity, dominance, cancellation and ordering of alternatives (see, e.g., Plous 2001: 81-82).

\(^{11}\) There are two main types of axiomatization of the expected utility theory: while the first one was firstly proposed by von Neumann and Morgenstern (1947) and is based on an objective notion of probability, the second kind of axiomatization is part of the so-called Bayesian decision theory (Savage 1954).
formal structure, the validity a deductive argument is independent of the matter or meaning of the premises and a deductively valid argument is valid in every context. That has two important consequences for the evaluative question. Firstly, the universality of the norms implies that, when faced with the same reasoning problem, any rational subject must draw the same conclusion, that is, if they are given the same premises, good reasoning can only lead to one correct conclusion.\(^{12}\) Secondly, if the same norms are applicable in every context, then the researchers involved in the rationality debate need not debate over which norms should be applied when assessing reasoning performances. Indeed, when different subjects dealing with the same reasoning problem draw different conclusions, that must be either because (a) they rely on different information or assumptions or because (b) at least one of them is not reasoning in accordance with the appropriate normative principle. There are no other options, because a rational conclusion will always follow with necessity from a given set of premises.

### 2.3 Deontologism and normativity

It is generally agreed upon that people have to satisfy many basic requirements in order to count as rational. For example, according to the *standard picture*, a rational subject should avoid drawing illogical inferences from information given, have consistent probability assignments, maximise her expected utility in each choice, and the likes. Stating that to reason rationally is to reason in accordance with a set of normative principles, dictated by an accepted normative model of reasoning, the *standard picture* assumes that each rational performance is directed at meeting one or more norms of rationality. However, a question

\(^{12}\) As is pointed out by Harold Brown (1990: 17), when people reason from some premises to a conclusion in accordance with a definite set of norms, they free themselves from the arbitrariness characterizing fallacious reasoning and pseudo-reasoning.
that needs to be addressed is how these norms have to be characterized.\textsuperscript{13} Theorems, axioms and rules of the formal theories of reasoning are not themselves normative. As is noted by Samuels and his colleagues (2004: 164), axioms and theorems of standard probability theory do not imply directly that we should reason in accordance with them. Indeed, such axioms and theorems describe truths about logic and probability: the axiom $0 \leq P(a) \leq 1$, for example, states that the probability of an event $a$ corresponds to a number that ranges from 0 to 1. In contrast, the norms of rationality are taken to be not mere descriptions of states of affairs: they express “ought” claims. More specifically, as characterized by Renée Elio (2002b: 4), a norm of rationality specifies usually “[…] how something ought to be, based on \textit{a priori} considerations that, if followed, yield ‘success’ on some dimension”. These norms are rules or principles that specify patterns of behaviour which are required independently of any social or legal institution and regardless of what people do. So, they do not have any descriptive ambition: these norms simply state what a subject ought to believe or do in order to count as rational. But what form should the utterances that express such norms take? How are the norms of rationality to be characterized? Looking at the philosophical literature on rationality, there is no real agreement over their characterization.\textsuperscript{14} Many philosophers and psychologists assume that these norms are represented in a sentence-like format, regulated by the formalism of a deontic logic (see Sripada & Stich 2006: 292). For example, a possible formulation of the norm of rationality concerning the \textit{modus ponens} is the following:

\begin{verbatim}
(MP) For any subject $x$, rationality requires that if $x$ believes that $p$ and $x$ believes that if $p$ then $q$, then $x$ ought to believe $q$. (adapted from Stein 1996: 6)
\end{verbatim}

\textsuperscript{13} It is not my aim here to produce a full list of the norms of rationality, which indeed would be too difficult a task. I consider here only some significant normative principles of rationality relevant for my discussion.

\textsuperscript{14} For example, both in the philosophy of action and in epistemology one can find very different characterizations of the norms of rationality and of their normative force (see, e.g., respectively Broome 1999; Engel 2008).
One should then substitute the appropriate elements for the schematic letters: the name of an individual for “x”, a sentence for “p” and “q” and so on. However, such a formulation specifies a single norm of rationality. What usually happens in the rationality debate is that a researcher wants to draw a more general conclusion about human rationality. Indeed, researchers involved in the rationality debate are not only interested in the ability of the human subjects to conform to certain norms of rationality but, rather, want to answer to a more general and complex question such as “Are humans rational?”. And that question calls for a characterization of what I am going to call the rationality condition for individuals.

2.4 The rationality condition for individuals

Answering the classical question “Are humans rational?” requires something like a general condition for the rationality of individuals. Indeed, every deontological account of rationality assumes implicitly a definite general requirement, which connects an individual’s capacity to reason, implicitly or explicitly, in accordance with the relevant norms with her rationality. Even if they assume the same norms of rationality, deontologists may develop very different characterizations of this requirement, which states either necessary and sufficient conditions or is weakened in some degree. Here I characterize the condition for rationality assumed by the supporters of the standard picture of rationality.16

When we hold that a requirement is a necessary condition for attributing rationality, we may characterize it as:

15 The rationality condition can be interpreted as requiring that a rational subject be able to recognize in some way both the norms of rationality and when she falls short of them. If the requirement is formulated in this way, rationality should be assumed to be strongly related to consciousness. However, even acting appropriately for, or reasoning in accordance with, a norm of rationality is not sufficient evidence to be attributed the possession of that norm.

16 In the next two chapters, I will characterize the conditions for rationality based respectively on a non-standard deontological conception of rationality and on a consequentialist one (see Chapter 2, Section 3.1; Chapter 3, Section 1.1).
(R1) Given a definite set \( N \) of norms of rationality, a subject \( S \) qualifies as rational if and only if \( S \)'s behaviour conforms to \( N \).

In (R1) \( S \) is fully rational if and only if she conforms to all the norms of rationality belonging to \( N \). That is, \( S \)'s being fully rational and her conforming to any norm of rationality (belonging to \( N \)) are logically connected. If \( S \) is to be fully rational, she has to conform to each particular norm of rationality in \( N \). (R1) seems to be a very strong requirement for the rationality of individuals.

For a long time, those who have been involved in the rationality debate have seemed to be committed to (R1). It is noteworthy that (R1) does not have a real normative force: it expresses a general condition on the rationality of an individual. The condition which it puts forward merely states when a human subject can qualify as rational but does not provide any recommendation about what she ought to do. Indeed, the normative force of this requirement depends on its content, that is, the norms of rationality which belong to \( N \).

As seen above, most researchers have identified the set of norms belonging to \( N \) with those stated by the standard picture of rationality. But do humans habitually conform to these norms?

Classical norms of rationality such as having logically consistent beliefs and consistent probability assignments, behaving in order to maximize one’s expected utility, taking account of all relevant information in arriving at conclusions, intuitively seem to pose too high requirements to ordinary people. As a matter of fact, people often make decisions by means of which they gain no benefits and also utter contradictory claims or fail to infer conclusions: assuming that (R1) expresses the appropriate rationality condition for individuals and that it is possible to demonstrate that people’s reasoning performances systematic fall short of the classical normative standards of rationality, are we allowed to conclude that humans are irrational after all?
At first sight, this question sounds odd: how can an educated group of human beings, such as psychologists or philosophers, assess human rationality without at the same time drawing a conclusion about their own rationality and so about their own capacity to draw a conclusion about that? Considering the latter question as inappropriate, some philosophers and psychologists have indeed given a positive answer to the previous question on the basis of a series of empirical findings on human reasoning. As said above, their claims have given rise to a controversy, which is known as the rationality debate. The problem is that these negative claims about human rationality have been so abundantly discussed that it is difficult to reconstruct what the experimental evidence really shows. In the next section, I will try to get clear this controversial issue.

3. The rationality debate

In this section, after introducing three well-known experiments on human reasoning, I discuss the two main positions involved in the rationality debate, i.e., the Rationality and Irrationality Theses. My aim is to make explicit both their different assumptions and aims, and also their similarities. Then, I conclude the section by examining two classical objections proposed by the supporters of the Rationality thesis which are concerned with experimental studies’ implications for the rationality question.

3.1 The psychology of reasoning: three experimental cases

Since the early seventies, some cognitive psychologists have begun to investigate human reasoning empirically. In order to do that, they have devised a series of reasoning tasks in which experimental subjects were asked to solve specific types of reasoning problems. These tasks include, for example, assessing the logical validity of an argument, deciding
what evidence one needs to test a conditional rule, estimating the posterior probability of a hypothesis on the basis of the evidence provided and so on. The right answers of all these tasks are supposed to be specified by the normative principles set by the standard picture of rationality.

After having collected a relevant set of data about subjects’ reasoning performances, experimenters evaluate the nature and quality of these performances against the “right” answer, thus providing an evaluation of the quality of human reasoning (see Chater & Oaksford 2001: 194). As representatives of these researches, I present three well-known reasoning experiments which, according to their creators, have provided evidence that human beings are not able to reason in accordance with certain basic norms of rationality. These experiments are part of two different research programs. One is the research program based on Wason’s selection task which tests people’s performances on conditional reasoning. The other is the “heuristics and biases” research program, which has begun with the works of Daniel Kahneman and Amos Tversky and is concerned with probabilistic reasoning and decision making.17

3.1.1 The Wason Selection Task

In the sixties, the psychologist Peter Wason devised a reasoning problem that has been known as the selection task, which he used in a series of simple psychological experiments (Wason 1966; 1968). This reasoning task was designed as one of hypothesis-testing, but calls for deductive reasoning based on a strictly logical interpretation of conditionals. The classical indicative version of this task is the following (adapted from the description of the task given by Wason & Johnson-Laird 1972: 172-173):

17 I take these tasks into account not just because they are famous examples of empirical studies on human reasoning but because these experiments have been largely discussed by analytic philosophers and cognitive psychologists.
Four cards are in front of you. Each of these cards has a letter on one of its sides and a number on its other side. The first two cards show the side with the letter and the other two that with the number.

There is then a rule which applies only to the four cards:

If a card has a vowel on one side, then it has an even number on the other side.

Your task is to name those cards, and only those cards, which need to be turned over in order to determine whether the rule is true or false.

In order to solve the task correctly, participants in the experiment should only turn over the cards that are “potentially logically incompatible” (to borrow the expression used by Chater & Oaksford 2001: 195) with the conditional statement. If the analysis of the conditional as material conditional is accepted, the only situation logically incompatible with the conditional statement \( \text{if } p \text{ then } q \), which may be observed, consists in a card with \( p \) on one side and \( \text{not-} q \) on the other.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>If ( p ) then ( q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

Figure 1. The truth-table of the material conditional.
So, the correct response is to turn over two cards, that is, those showing the vowel \( p \) and respectively the odd number \( \text{not-}q \). Indeed, if by turning the card showing the vowel \( p \) the subject finds an odd number \( \text{not-}q \), then the conditional statement is falsified. Likewise, if the card showing the odd number \( \text{not-}q \) is turned over, the presence of a vowel \( p \) on the other side will falsify the conditional statement. If a subject turns over the card showing either the consonant \( \text{not-}p \) or the even number \( q \), that can not provide falsifying instances: whatever these cards may have on their hidden sides cannot falsify the conditional statement. That can be formalized as a normative principle in the following way:

\[(CT) \text{ For any subject } x, \text{ rationality requires that, in order to test the truth of a conditional statement, } x \text{ ought to consider both cases where the antecedent is true so as to check whether the consequent is true and cases where the consequent is false so as to check whether the antecedent is false. (adapted from Stein 1996: 81)}\]

Most participants to the Wason’s experiment failed to conform to CT. Indeed, only 4% reached the correct response in the first experiments summarised by Wason and Johnson-Laird (1972: 182): 46% of the subjects selected the A-card and the 2-card and 33% of them the A-card alone (the remaining 17% of the subjects chose other combinations). The same experiment was replicated more than once and the results were almost the same (Wason 1968; 1969; Wason & Johnson-Laird 1970; 1972). Moreover, some experimental subjects continued to make the same error even after being subjected to “therapeutic” procedures aimed at making evident to them the need to check for falsifying instances (Wason 1969).\(^\text{18}\)

\(^{18}\) Subsequent studies have shown that subjects’ selections vary according to how the task is formulated (for a short introduction to such studies, see Chapter 5, Section 3).
3.1.2 The Linda Problem

A well-known example of an experiment that tests people’s ability for probabilistic reasoning is the so-called “Linda Problem”. This experiment was proposed for the first time by Amos Tversky and Daniel Kahneman (1982; 1983). In its standard version, subjects were presented with the following task (reported by Tversky & Kahneman 1982: 92):

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.

(a) Linda is a teacher in elementary school.
(b) Linda works in a bookstore and takes Yoga classes.
(c) Linda is active in the feminist movement.
(d) Linda is a psychiatric social worker.
(e) Linda is a member of the League of Women Voters.
(f) Linda is a bank teller.
(g) Linda is an insurance salesperson.
(h) Linda is a bank teller and is active in the feminist movement.

Tversky and Kahneman presented the task to a group of subjects who had never attended any course in statistics or probability. Most of them (about 85%) ranked the conjunction (h) as more probable than its conjunct (f), violating the conjunction rule of probability theory, according to which the probability of a conjunction cannot be greater than the probability of its constituents. Indeed, if Linda is a bank teller and is active in the feminist movement (as presented in (h)), then, necessarily, she is a bank teller (as presented in (f)). While whenever the event corresponding to (h) takes place, necessarily, the event corresponding to (f) takes place as well, it may be the case that the latter takes place without (h) taking
place: in other words, she may be a bank teller without being active in the feminist movement. As a consequence, it is not the case that the event corresponding to (h) may be more probable to take place than (f). So, most reasoning performances in this experiment do not seem to conform to the following basic normative principle of rationality:

(CP) For any subject $x$, rationality requires that $x$ ought to assign a greater (or equal) degree of probability to the occurrence of a particular event than one does to the conjoint occurrence of that event and another distinct event. (adapted from Stein 1996: 6)

Tversky and Kahneman (1983) demonstrated also that the error does not depend on the specific content, as it is performed in experiments where subjects were asked to estimate probabilities of events belonging to different domains.

3.1.3 The Asian Disease Problem

This task concerns a decision problem devised by Tversky and Kahneman (1981: 453). In this experiment, participants were presented with two problems (see below) and then requested, in each, to make a choice between two competing programs:

**Problem 1:** Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that all 600 people will be saved and, a two-thirds probability that no people will be saved. Which of the two programs would you favour?

**Problem 2:** Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat
the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. Which of the two programs would you favour?

When examined side by side, the two pairs of programs are equivalent as to their results. If you consider the programs A and C, you can see that, while both referring to 600 lives, saving 200 lives is equivalent to losing 400 live (Program A) and losing 400 lives is equivalent to saving 200 lives (Program C). Likewise, although differently described, the outcomes of the programs B and D coincide. However, when they answer to the two problems separately, most participants choose the programs A (72%) and respectively D (78%). Insofar as the results of the programs C and D are equivalent to those of the programs A and B in each pair of problems, those who choose A and D or B and C are behaving in violation of the normative principle of rationality derived by the axiom of invariance, which claims that “preference order between prospects should not depend on the manner in which they are described” (Kahneman & Tversky 1984: 343):

(AI) For any subject \( x \), rationality requires that if \( x \) prefers \( a \) to \( b \), she always ought to prefer \( a \) to \( b \), regardless of how they are presented.\(^{19}\)

3.1.4 Simplicity vs. complexity

There are a number of similar experimental results which seem to show that people’s reasoning systematically falls short of classical standards of rationality. As is well summarized by Keith Stanovich (1999: 1-2),

\(^{19}\) As is noted by Keys and Schwartz (2007: 163), the axiom of invariance is considered so obvious that “many accounts that attempt to formalize the rules of rational decision making use it implicitly, without stating it explicitly”.

28
people assess probabilities incorrectly, they display confirmation bias, they test
hypothesis inefficiently, they violate the axioms of utility theory, they do not
properly calibrate degrees of beliefs, they overproject their own opinions onto
others, they allow prior knowledge to become implicated in deductive reasoning,
they systematically underweight information about non-occurrence when
evaluating covariation, and they display numerous other information processing
biases.

Surely the reason why so much research interest has focused on the reasoning tasks just
illustrated is precisely because despite their apparent simplicity, most intelligent subjects
fail to solve them. No one would be interested in errors on a task of obvious computational
complexity. In the case of the selection task, for instance, participants are faced with only
four cards and in order to solve it correctly they ought to consider only two logical
possibilities concerning the hidden sides of each. Many other reasoning tasks employed by
psychologists involve relatively simple principles of reasoning and structures. In this sense,
systematic violations cannot be explained by assuming that these reasoning tasks are too
difficult for human subjects. It has been argued, instead, that such systematic violations
seem to be due to the ways in which people reason and make decisions. That is why many
analytic philosophers and cognitive psychologists have acknowledged the relevance of
those experimental data for examining to which extent human reasoning competence meets
the normative standards of rationality and so to the extent to which humans can be really
regarded as rational.

3.2 Rationality vs. Irrationality Theses

But what do such experimental findings on human reasoning show? There has been a
considerable debate over this question as the interpretation of these empirical results and
the standards used to assess subjects’ performances on reasoning tasks are not straightforward matters. Looking at the vast literature concerning this debate (see, e.g., Jungermann 1986; Stein 1996), we may distinguish between two general positions, the Rationality thesis and the Irrationality thesis, both of which assume the *standard picture* and its deontological conception of rationality. By examining the empirical results on human reasoning, their supporters have focused on whether and to what extent the principles that are part of human reasoning competence comply with those belonging to the *standard picture of rationality*. Human reasoning competence has been regarded as a set of principles of reasoning which are cognitively represented: on the one hand, supporters of the Rationality thesis argue that the principles being in human reasoning competence correspond to those that people ought to follow and, on the other hand, supporters of the Irrationality thesis claim that the principles being in human reasoning competence are different from those they ought to follow in order to count as rational (Stein 1996: 10).

According to this view, when you assess an instance of reasoning, you are examining whether that piece of reasoning is in accordance with the relevant normative principle of reasoning in that situation. If that piece of reasoning is not in accordance with what is dictated by the relevant normative principle, that may be either because that principle is not part of the human reasoning competence or because that piece of reasoning results from the interaction between the correct application of the relevant normative principle and some interfering factors that are extrinsic to human reasoning competence. The characterization

---

20 This does not itself mean that humans follow the normative principles of rationality consciously; cognitive processes underlying reasoning just may conform to them at a sub-personal level (not consciously) and fairly reliably.

21 Clearly, this is an oversimplification. One cannot simply decide whether or not people possess a particular principle of reasoning by assessing their performance on a single type of problem which requires the use of that principle. In order to draw appropriate conclusions it is necessary to investigate human reasoning when a number of features of the form and content the problems are varied.

22 This distinction between competence and performance corresponds to that employed by Chomsky in the field of linguistics (see, e.g., Chomsky 1965). Roughly speaking, according to Chomsky, competence amounts to people’s underlying knowledge of language, and performance refers to what people actually do when these
of human reasoning competence becomes the key to rationality and consequently the question of human rationality is regarded as being concerned with whether or not the relevant normative principles of reasoning are part of the human reasoning competence. As is pointed out by Robert Hanna (2006: 124-125), the supporters of the Rationality and Irrationality theses agree in all their premises and differ only in their conclusions. Given all the empirical evidence of systematic failures in human reasoning, and assuming

(a) a normative characterization of rationality;

(b) that normative principles of rationality stem from deductive logic, standard probability theory and expected utility theory;

(c) a deontologist conception of rationality;

(d) a rationality condition for individuals such as (R1);

then

(i) if it is true that people do not systematically reason in accordance with the normative principles of rationality, then it follows that they are fundamentally irrational;

(ii) if it is possible to demonstrate both that, when people’s reasoning performances diverge from the normative standards of rationality, their divergences are attributable to factors extrinsic to human reasoning competence and that, when their reasoning performances accord with what is stated by the appropriate normative principles, performances are the result of their reasoning competence, then it follows that people can be considered as rational after all.  

That is what Robert Hanna (2006: 125) calls “the crux of the rationality debate”. More appropriately, I hold, this argument articulates in a clear way the main assumptions
underlying the evaluative framework to which both the supporters of the Rationality thesis and those of the Irrationality thesis are committed. As we have seen above, starting from the same evaluative framework and the same normative standards of rationality, they conclude in a significantly different way (see, e.g., Jungermann 1986; Samuels et al. 2004; Stein 1996). On the one hand, according to the supporters of the Irrationality thesis, what empirical studies on human reasoning show is that people lack the appropriate principles of rationality in their reasoning competence. In their view, subjects’ reasoning performances systematically fall short of the appropriate normative principles of rationality because people tend to rely on heuristics “which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors” (Kahneman & Tversky 1973: 48). On this basis, the supporters of the Irrationality Thesis conclude that human reasoning competence should be considered as at least normatively defective since it does not possess or possesses only in a limited way the appropriate principles of reasoning. That is, they maintain that the systematic violations of basic norms of rationality are attributable exclusively to people’s faulty reasoning competence. On the other hand, according to the supporters of the Rationality thesis, their opponents claim far too much generality for their empirical findings, so that the picture of human irrationality they provide is misleading. According to them, the empirical evidence against people’s ability to reason rationally is not so compelling: the errors that subjects make are the result of the interaction between cognitive processes underlying reasoning and either other cognitive processes or contextual variables or both. Even if the cognitive processes underlying reasoning conform to the appropriate principles of reasoning, the activity of those processes can sometimes give rise to performance errors, that is, our reasoning performances do not always reflect the principles which constitute our reasoning competence. Most of the allegedly inappropriate responses in reasoning tasks should not be considered as reliable indicators of the fact that our
reasoning competence lacking the normatively appropriate principles of reasoning. On this view, it is possible for someone to be rational by having a normatively appropriate reasoning competence, yet fail to reason appropriately on particular occasions. Moreover, as we will see, it has been demonstrated that there are many reasoning problems on which people’s performances do not really deviate from the relevant normative standards of rationality. As a consequence, contrary to what is stated by the supporters of the Irrationality thesis, the Rationality thesis claims that “all normal humans have the capacity to reason in accordance with the normative principles of reasoning” (Stein 1996: 11).

3.3 Systematic violations or occasional mistakes?
Consider now how supporters of the Rationality thesis explain systematic violations in reasoning tasks. They usually maintain that not all errors in reasoning tasks should be taken to indicate irrationality: failing to act in a way that satisfies the norms of rationality is not always a sign of irrationality, “but rather is just due to having made a mistake” (Stein 1996: 8). But what is the difference between the kind of errors involved in being irrational and occasional mistakes that do not involve irrationality? As is explained by Stein, supporters of the Rationality thesis maintain that “[b]ehind the notion of a mere mistake is the idea that making a mistake involves a momentary lapse, a divergence from some typical behaviour” (Stein 1996: 8). Opposite to that, what supporters of the Irrationality thesis claim is that “behaviour due to irrationality connote a systematic divergence from the norm” (Stein 1996: 8). As a consequence, both supporters of the Irrationality thesis and of the Rationality thesis assume the distinction between occasional mistakes and systematic violations when they use the results of empirical studies on human reasoning as evidence in favour of their theses. According to the supporters of the Irrationality thesis, when researchers find a systematic violation in a reasoning task, we must conclude that people do not possess the
principle of reasoning to which they should have conformed in that task. In the same vein, the supporters of the Rationality thesis implicitly assume such a distinction in order to deny that empirical evidence on human reasoning is relevant in deciding whether or not their thesis is true (Stein 1996: 8). The latter thesis could be saved only if its supporters can show that people have the psychological capacity to reason appropriately, but on particular occasions do not display that capacity. Differently, according to the supporters of the Irrationality thesis, if humans really had the capacity to solve a certain type of reasoning problem, they would be able to solve it independently of the context or subject matter. In particular, in the case of successful performances on particular revised versions of a reasoning task, the supporters of the Irrationality thesis argue that the added elements trigger heuristics making it possible for subjects to conclude the task successfully. From the point of view of the Irrationality thesis, the experimental findings on human reasoning have “bleak implications for human rationality” (Nisbett & Borgida 1975: 935): what they indicate is that much human behaviour is produced by unreliable heuristics rather than rational processes. Indeed, if the heuristics had reliably yielded rational beliefs or decisions, they would have been acceptable. The problem here is that supporters of the Irrationality thesis hold that heuristics do not do that. So, they hold that we have no general psychological capacity to reason systematically in accordance with normative principles of rationality, and conclude that humans are fundamentally irrational.

3.4 Heuristics and biases
These considerations require a clarification about the notion of a heuristic. In the rationality debate, that notion is used by the researchers belonging to the so-called “heuristics and biases” approach. The two key concepts just mentioned (heuristic and bias) are at the core
of this approach. Even if heuristics have never been precisely defined in it,\textsuperscript{24} they are considered as specific procedures that yield reliable inferences in some specific range of contexts or environments (see Gilcovich \textit{et al.} 2002: 1). Heuristics are usually described as cognitive shortcuts or rules of thumb for reasoning. As Elizabeth Hurley (2005: 507) has pointed out, a controversial issue is “how reliable the use of heuristics must be, across what range of variation in possible environments, in order to be compatible with rational behavior”. As we have just said, according to the supporters of the Irrationality thesis, since heuristics do not lead reliably to rational behaviour, they are not rational processes.\textsuperscript{25} In order to support this claim, they compare heuristic process to algorithmic ones holding that only the latter lead reliably to good outcomes (if used correctly).\textsuperscript{26} Relying on a heuristic process always involves both advantages and risks, as they may lead people by a short cut (compared to algorithmic processing) to the goal they seek: on the one hand, a heuristic process can lead to good results most of the time, even if not all of the time, and, on the other hand, it works without expending too much time and effort (compared to algorithmic processing). Heuristics allow people to make rapid inferences and decisions without having to take into account all relevant data or to perform complex computations as they “reduce the complexity of judgement tasks, to make them tractable for the kind of mind that people happen to have” (Kahneman \textit{et al.} 1982: xii). However, overreliance on heuristics can lead to systematic errors in reasoning, which are known as biases.\textsuperscript{27} Biases are deviations from a

\textsuperscript{24} For example, according to Gerd Gigerenzer (1996), heuristics, as the heuristics and biases approach defines them, are so vaguely specified that they are not so much informative, and even testable.

\textsuperscript{25} Contrary to that, in the next two chapters, we will see that many researchers involved in the rationality debate have argued for the rationality of the heuristic processes (see Chapter 2, Section 2; Chapter 3, Section 2 and 3).

\textsuperscript{26} More specifically, when applied to a problem, an algorithmic process leads to the correct solution in a finite number of steps (if employed correctly), such as in the case of the truth-table test for validity. Such a procedure has mainly two advantages: (i) it is mechanically applicable; (ii) it consists of a decision procedure which indicates whether a given argument is valid or invalid. However, such decision procedures can be applied only to relatively simple problems (see Brown 1990: 20-21)

\textsuperscript{27} According to Goldman (1999: 230), a distinction can be made between hot and cold biases. Hot biases are determined by desires, emotions, and preferences that do not allow people to look at how things actually are. Cold biases are instead due to the inherent characteristics of the human mind.
normative principle which are caused by overreliance on heuristics, but also due to
cognitive limitations, perceptual organizations and the like. In this sense, as is pointed out
by Keren and Teigen (2004: 91-92), biases are conceived as effects, but often, as for
example in psychological studies based on logical tasks, systematic failures in deductive
reasoning are attributable to more general biases, such as the “confirmation bias”, that is, a
tendency to test hypotheses by checking for empirical evidence that confirms rather than
falsifies them. That is worth remembering as biases can be conceived as effect (to be
explained by heuristics) and as causes of systematic deviations. Take as examples the Linda
Problem and the Asian Disease Problem. In the first case, Kahneman and Tversky (1973:
48) hold that people rely on the so-called representativeness heuristic, according to which,
“[people] select or order outcomes by the degree to which the outcomes represent the
essential features of the evidence. In many situations, representative outcomes are indeed
more likely than others”. In the Linda Problem, while “being an active feminist” well
summarizes the description of Linda, being a bank teller does not. As a consequence, when
faced with a conjunction consisting of a representative characteristic (feminism) and a non-
representative one (Linda’s being a bank teller) as in the case of “Linda is a bank teller and
is active in the feminist movement” and a statement consisting only of the non-
representative constituent of the conjunction as in the case of “Linda is a bank teller”,
people are naturally inclined to judge more probable the former than the latter, although
logically it ought to be the opposite.\textsuperscript{28} In the Asian Disease Problem, reasoning
performances display a different bias caused by the so-called framing effect. Here, the
problem is that the way things are described influences how people reason about them.

\textsuperscript{28} Several criticisms have been levelled against the conclusion reached by Tversky and Kahneman. Most
prominently we can divide those criticisms into two categories: on the one side, some scholars have proposed
a different explanation of subjects’ performances based on a frequentist interpretation of probability (Fiedler
1988; Gigerenzer 1991; 1995; Hertwig & Gigerenzer 1998) and, on the other hand, others have proposed a
critical assessment of the task from a pragmatic point of view (Dulany & Hilton 1991; Politzer & Noveck
With reference to the Asian Disease Problem, Kahneman and Tversky (1981) hold that people have a tendency to be risk taking when considering losses (people who die) but to be more risk averse when considering gains (people saved) because their decisions are strongly influenced by how potential outcomes are represented in the problem. So, what the results of these experiments indicate is that it is extremely important whether something is framed as a gain or as a loss because different ways of framing a problem could lead to counter-normative selections.

This brief presentation raises a question: if heuristics are not so reliable, why should people rely on them? The problem is that people rely on heuristics frequently, but usually without being aware of it. Kahneman and Tversky have compared them to “visual illusions”, dubbing them “cognitive illusions”. Cognitive illusions, as well as visual illusions, consist of situations in which people spontaneously see or understand things differently from how they actually are. Just as, when visual illusions are pointed out, people usually continue to see the situation in the misleading manner, so happens with cognitive illusions too. I hold that this point explains well what Kahneman and Tversky, as well as the other supporters of the Irrationality thesis, claim when they argue that humans are to some extent irrational: irrationality is something intrinsic to the human cognitive system, even if humans are not conscious of that. That, I think, is the core of the Irrationality thesis.

3.5 Two general arguments in defence of human rationality

Contrary to the claims of the researchers involved in the heuristics and biases approach, the supporters of the Rationality thesis maintain that any psychological evidence that seems to show systematic failures in human reasoning can be properly re-interpreted in order to defend the rationality of human beings. As representatives of these attempts, I introduce and then criticize two well-known arguments (one philosophical, the other closer to
psychology) which have been discussed at great length among the researchers involved in the rationality debate.

3.5.1 The necessary presumption of rationality

According to a received view widespread among philosophers of language and mind, assuming that people are rational is a necessary condition to explain and predict human behaviour with good approximation. Such a presumption of rationality is held to be so strongly entrenched in our ordinary life and not be, therefore, easily questionable. Those who defend human rationality rely in particular upon three arguments based on the presumption of rationality. Before all, they refer to the works of W.V. Quine. In *Word and Object* (1960), he proposed the so-called Principle of Charity in the attempt to explain how a field linguist can appropriately translate a totally unknown language into her home language. According to this principle, in order to understand the utterances of a speaker of an unknown language the field linguist should translate the native speaker’s utterances so as to avoid interpreting them as absurd or contradictory. As Quine stated,

(i) “interlocutor’s silliness, beyond a certain point, is less likely than bad translation”

(Quine 1960: 59), where silliness has to be evaluated from the translator’s perspective (see Quine 1960: 69);

(ii) “fair translation preserves logical laws” (Quine 1960: 59).

Put another way, Quine suggested that we should normally consider an apparently nonsensical translation as due to a mere mistake of the translator rather than irrationality on the part of the speaker.\(^{29}\) These considerations have been developed in both Donald Davidson’s theory of radical interpretation and Daniel Dennett’s theory of intentional

\(^{29}\) Quine’s claims about radical translation are, however, ambiguous. As is pointed out by Thagard and Nisbett (1983: 252), it is not clear “how stringently we are to take the injunction to ‘avoid’ imputing irrationality”.

38
stance. Their theories are usually considered good arguments against the interpretations of
the empirical evidence about human reasoning which are supported by the proponents of
the Irrationality thesis. Indeed, according to such theories, in order to be able to understand
your interlocutor, you have to assume that she will be on occasion rational: no-one would
be able to understand a non-rational individual. On the one hand, Donald Davidson argues
for a stronger form of the Principle of Charity, according to which it is a necessary
condition for the interpretation of someone’s speech and behaviour to assume that she is by
and large rational and most of her beliefs are true (Davidson 1982: 302-303; 1984: 183-
198). While Quine uses the Principle of Charity as a methodological principle in
translation, Davidson considers it as an a priori principle governing both the ascription of
propositional attitudes and the interpretation of behaviour (Davidson 1984: 197). He claims
that “if we are intelligently to attribute attitudes and beliefs, or usefully to describe motions
as behaviour, then we are committed to finding in the pattern of behaviour, belief, and
desire a large degree of rationality and consistency” (Davidson 1982: 237). In this sense,
successful interpretations necessarily require that the speaker interpreted is imputed with a
high degree of rationality. On the other hand, Daniel Dennett argues that the attribution of
rationality is a precondition of our ability to attribute intentionality to other beings (humans
or not): “the assumption that something is an intentional system is the assumption that it is
rational” (Dennett 1981: 10). When one adopts what Dennett calls “the intentional stance”
toward something, (i) she first attributes an intentional attitude to it and, then, on the basis
of this attribution, (ii) she should be able to predict its behaviour by determining what
actions are rationally required to be accomplished on the basis of the attitudes which has
been attributed to it (Dennett 1981: 6-9).

What implications do these arguments have for the rationality debate? The main claim
being made by these arguments is that we cannot understand speech, interpret behaviour, or
attribute intentional attitudes without also ascribing a high degree of rationality to other’s thoughts and behaviour. The supporters of these arguments conclude from this that it is conceptually impossible that people’s reasoning performances systematically fall short of the classical normative standards of rationality. As is pointed out by Pascal Engel (1989: 306), they assume that, when assessing people’s rationality, one must assume that some principles of rationality apply *a priori* to their behaviour, and such principles cannot be examined empirically, because any evaluation of people’s rationality can be realized only in the light of these *a priori* principles. According to such a view, “normativity does not bear so much on the evaluation or on the prescription of how people should reason, but on the explanation or the prediction of events” (Engel 1989: 306). However, such a sense of normativity is very different from that assumed by the researchers involved in the rationality debate. In their works, what a normative model of rationality provides are norms, standards or criteria that allow assessing an instance of reasoning as correct or incorrect. According to the former view, put forth by Davidson and Dennett and supported by Engel, when discussing the rationality of people’s behaviour we should presuppose the ascription to them of certain basic norms of rationality governing their reasoning and decisions: their behaviour can be explained and predicted only on the basis of such a presumption of rationality. According to the latter view, supported by the researchers involved in the rationality debate, when studying human reasoning we cannot presuppose that some basic norms of rationality apply *a priori* to people’s behaviour because any claim about people’s rationality has to be empirically evaluated: we cannot attribute the ability to follow certain basic principles of rationality to people regardless of whether they actually conform to them. So, you can find here two different ways of conceiving normativity which work at two different levels.
What I find questionable in the arguments put forth by Davidson and Dennett is that, although both their accounts clearly presuppose that, in one form or another, the understanding and predictability of individual’s speech and behaviour depends on the assumption that she will speak and behave in conformity to a set of *a priori* principles of rationality, they have failed to provide a clear characterization of these principles. As is argued by William Child (1994: 56-89), Davidson’s argument can be interpreted as supporting what Child calls the thesis of the uncodifiability of rationality, according to which it is impossible to formulate normative principles determining what it is rational to do or believe in a given situation. In the same vein, Dennett argues that, in interpreting another’s speech or behaviour, we have to rely on our pre-theoretical concept of rationality, which is based on our shared intuitions about what makes sense in another’s speech or behaviour (Dennett 1987: 98). If so, in order to defend human rationality, these arguments make the concept of rationality pointless. Indeed, by accepting these characterizations, rationality lacks what any normative concept needs, that is, association with a set of definite norms. Accordingly, both Davidson and Dennett’s arguments do not seem to show that there really are *a priori* principles against which any evaluation of people’s rationality should be framed.

3.5.2 The alternative task interpretation argument

According to this second argument, while the experimenter may be assessing the subject’s answer as incorrect, the subject might have interpreted the reasoning problem she is presented with in a way different from the experimenter’s, and therefore have performed a
different task. So, the subject may have given the right answer to a different question (for a
discussion see Stanovich 1999: 98-141). This can happen because reasoning problems
presented to subjects are often vaguely posed. When this occurs without the experimenter’s
being aware of it, the subjects’ reasoning performance may be assessed against an
inappropriate normative principle.

This is a central problem in the study of human reasoning. As Nickerson (2008: 49)
remarks, a reasoning task devised to examine whether the subjects do A or B, where A and
B are regarded as the entire set of possibilities, may provide misleading results if what they
actually do is something different from A or B. Moreover, if, assuming that people
habitually approach a particular reasoning problem in a specified way, an experimenter
devises a reasoning task intended to show whether they solve such a problem in a
normatively appropriate way, experimental results will not be very informative if subjects
approach the problem in a way that differs from what the experimenter assumed to be the
case. In such a case, subjects could have misinterpreted part of the information available in
the experimental context, and then reasoned rationally from the misinterpreted information.
However, it is generally open only to the experimenter to question the subjects’
interpretation of the task.

There has been considerable debate in the psychology of reasoning about whether
apparent faulty reasoning performances might be explained on the basis of what Keith
Stenning and Michiel van Lambalgen (2001: 285) have called “non-standard
interpretations”, that is, any interpretation which significantly differ from the one assumed
by the experimenter. Two psychologists have been the main proponents of this argument:
Mary Henle and Jan Smedslund (see, e.g., Henle 1962; Smedslund 1990). Although their
arguments differ in some respects, both of them maintain that we cannot consider other
people’s reasoning performances as irrational, because we should always assume that their
performances are a rational consequence of their momentary premises. Their “alternative task interpretation” argument could be spelled out in the following way:

- if $S$ understands reasoning problem $P$ as $P'$ and
- if the answer $A$ is the normatively appropriate response to $P'$, then $S$ reasons rationally when she answers $A$ to problem $P$;
- $S$ understands reasoning problem $P$ as $P'$ and answer $A$ is the normatively appropriate answer to $P'$;
- so, $S$ reasons rationally when she answers $A$ to $P'$.

But the conclusion of this argument cannot follow, unless we add a condition:

- $S$’s understanding of $P$ as $P'$ is legitimate.\textsuperscript{32}

On this view, the worry for the supporters of the alternative task interpretation argument should be to explain in which cases the subject’s task interpretation is legitimate. The burden of the proof is on them: is it legitimate for a subject to ignore the instructions, change the information given, and bring in extraneous information or knowledge? The supporters of this argument should provide (at least) something as a set of criteria, according to which we can distinguish whether or not the subject’s understanding of the reasoning problem is legitimate. As is pointed out by Howard Margolis, many of the alternative task interpretations “[…] are so bizarre – so far from what the very words in the instructions said – that they represent serious cognitive errors that deserve attention” (Margolis 1987: 20: as reported by Stanovich 1999: 189). According to Stanovich (1999: 189), while attempting to defend the rationality of human beings, the proponents of the alternative task interpretation argument transfer irrationality from people’s reasoning competence to the stage of problem interpretation (or problem framing). So, irrationality has not disappeared, it is simply transferred to another cognitive activity.

\textsuperscript{32}This argument is adapted from one put forth by Jonathan Adler (1999: 666).
4. The standard picture of rationality in question

By examining empirical evidence on human reasoning, the researchers involved in the rationality debate seem to face two alternatives: if people systematically fall short of the normative principles set by the *standard picture of rationality*, that may be either because they are really departing from its normative principles and so they can be considered as irrational, or because their reasoning performances are influenced by some factors that are extrinsic to human reasoning. However, there is a third option: calling into question the appropriateness of the normative model against which experimenters assess reasoning performances. Perhaps, this is the most telling argument to arise against the supporters of the Rationality and the Irrationality theses. Many researchers take the *standard picture of rationality* as an absolute standard of correctness when evaluating human reasoning. But they rarely provide an argument or an explanation of why its normative principles are the right ones. Contrary to that, I would like to introduce two strong arguments against the use of the *standard picture* as a standard of correctness. These arguments cast doubt on the universal applicability and necessity of the *standard picture’s* normative principles, that is, on what its supporters are its two main virtues.

4.1 Which normative model is the right one?

In most empirical studies on human reasoning, psychologists have assumed that every reasoning problem has only one correct and uncontroversial response dictated by the normative model of rationality to which they are committed. In particular, any claim about biases and errors in reasoning is made by assuming the only validity of the *standard picture of rationality* alone, which most researchers are committed to, and its correct application. In other words, they consider the *standard picture of rationality* as an absolute standard of correctness when evaluating human reasoning. Looking at the literature on reasoning
experiments, one can see that those who assume the standard picture of rationality do not provide any explicit reason why we should ground normative models of rationality on formal theories of reasoning such as deductive logic, standard probability theory, and expected utility theory; rather they seem to be simply taking it for granted that it is so. However, if you consider them in a loose way, logic, probability theory, and expected utility theory are not monolithic enterprises and hence that raises the question of which version of them gives the true normative model of rationality. As is noted by Samuels and his colleagues (2004: 160), in each of these disciplines “there is a wide range of formal theories which make incompatible claims, and it’s far from clear which of these theories are the ones from which normative principles of reasoning ought to be derived”. In particular, there are some formal theories of reasoning whose principles are more relevant to the ways people reason in their ordinary life than deductive logic, standard probability theory and expected utility theory do and they can explain many of the alleged faulty reasoning performances that result from empirical research. For example, with regard to probability theory, there are different interpretations of probability that account for our judgments about the probabilities of individual events and accordingly there are alternatives to the Bayesian account of probability as a normative touchstone against which to evaluate human reasoning (see, e.g., Nozick 1993). In the same vein, in the domain of logic, you can find plenty of non-classical logics, e.g., many-valued logics, paraconsistent logic, relevance logic (for an overview see, e.g., Jacquette 2006). For example, relevance logicians have proposed a logical system that claims to be able to solve some of the classical problems of deductive logic (e.g., Mares 2004). In particular, relevance logic provides a formal system which “allows only those deductions that, in addition to being truth-preserving, ensure the relevance of the premises of the argument to the conclusion” (Baghramian 2004: 329). According to its proponents, relevance logic reflects ordinary reasoning in a way that is
much more accurately than classical deductive logic. So, why should we prefer deductive logic, standard probability theory and expected utility theory as the formal theories of reasoning from which normative principles of rationality are derived? There do not seem to be compelling reasons or arguments according to which we should prefer them to any other competing theory. A supporter of the standard picture of rationality could appeal to its main virtues, that is, the necessity and the universality of its principles. But, as we will see soon, if we look better at these two alleged virtues, they seem to be a limit rather than an advantage when using the normative principles set by the standard picture for assessing reasoning performances.

4.2 Universality and necessity in rationality assessments

Suppose that, however, we accept the standard picture of rationality, there remains to be established whether the norms of rationality derived from deductive logic, standard probability theory, and expected utility theory can be appropriately applied to subjects’ reasoning performances in order to assess them. In the last twenty years or so, it has been widely agreed upon that such norms are of very limited utility when dealing with ordinary situations (where reasoning performances usually take place). Indeed, according to their critics, being assumed to be universal and necessary, axioms and principles derived from classical theories of reasoning (such as Bayes’s theorem, consistency, transitivity, etc.) are deemed to be by themselves indeterminate with respect to people’s reasoning performances in ordinary situations (Gigerenzer 1998; Harman 1986; 1995). Since they are abstract (content-free) and too formal, their critics conclude that the normative principles being part of the standard picture of rationality are not as good in evaluating human reasoning as the supporters of the Rationality and Irrationality theses maintain.

33 It is not my purpose here to argue about the merits and demerits of alternative interpretations of probability and of alternative systems of logic.
Consider classical deductive logic and the alleged principles of rationality which can be derived from it. As is argued by Harman (1995), while classical deductive logic is concerned with arguments, which consist of abstract relations among propositions, it is far from clear how a normative theory of rationality, that is, a theory of how people ought to reason, can be derived from its axioms and principles. What logicians call deductive validity involves only the relationship between the premises and the conclusion, so that it cannot be the case that when in a deductive argument all the premises are true a false conclusion occurs. But it does not say anything about how cognitive processes underlying deductive reasoning should work. Furthermore, classical deductive logic cannot determine which of the infinitely many deductively valid conclusions it is sensible for us to draw. From any set of premises, there is always an indefinite number of conclusions that can be validly drawn. But most of the conclusions are so trivial and irrelevant that no rational individual would draw them (see, e.g., Harman 1995: 18-21; 21-23). On this view, deductive logic cannot say anything about what conclusions are worth deriving in a given situation. That is why principles of rationality derived from deductive logic are of little help in assessing how people ought to reason in ordinary situations.

In sum, the key question here is whether the normative principles derived from the standard picture of rationality provide us with an appropriate normative background against which to assess people’s reasoning. There are strong limits in using the normative principles belonging to the standard picture of rationality to assess human reasoning, because there are only few cases in which they can be applied appropriately. Contrary to what is argued by the supporters of the Rationality and the Irrationality theses, the normative principles belonging to the standard picture can be applied to evaluate the quality of human reasoning only in situations in which it is possible to clearly identify which are the premises of the reasoning (and their logical form) and what is rationally
required to do with them. But that rarely happens in ordinary situations. That is why the norms of rationality belonging to the *standard picture of rationality* have a very limited range of application. So, what the supporters of the *standard picture of rationality* consider to be the two main reasons for adopting it when evaluating reasoning performances, that is, the universality and the necessity of its normative principles, become the real limits to its application to a great variety of reasoning cases. This does not mean that its normative principles fail to be significant in the study of human reasoning and rationality, but rather that what is at stake is the way those principles are applied to concrete situations. In other words, this objection concerns the application of the *standard picture of rationality* rather than the validity of its normative principles in the abstract.

### 5. Concluding remarks

In this chapter, I have examined what I have called the standard deontological approach to rationality assessment. I have distinguished the researchers committed to this approach in two groups on the basis of their claims regarding the extent to which the principles being in human reasoning competence correspond to those belonging to the *standard picture of rationality*, especially as shown by the experimental results on human reasoning. While this distinction can be very useful to understand what is at stake in the rationality debate, it is however an oversimplification. As a matter of fact, only few of the researchers mentioned above would argue either that people always conform to the normative principles of rationality or they never do. What I have wanted to show is that the scholars involved in the rationality debate, especially those who have supported the Irrationality thesis, have not paid much attention to the role that the normative standards of rationality plays in their interpretation of the empirical findings and of the fact that other normative standards may
be possible. Any of their claims about biases in reasoning presuppose the validity of the *standard picture of rationality* alone, which most researchers are committed to, and its correct application. But are its normative principles always appropriate for assessing reasoning performances? At the end of the section, we have seen two arguments providing a negative answer to this question. On the other side, those who defend human rationality seem to face with similar problems. They argue that failures in reasoning could be only attributed to factors extrinsic to human reasoning competence, such as working memory limitations, attention limitations, carelessness and so on. However, anyone can add new factors in this list so that any instance of what appears to be an error in reasoning performance can be easily explained away. As Nickerson points out, “inasmuch as any instance of faulty reasoning can be attributed to some determinant of performance other than competence, a faulty competence can never be established from observation of faulty performance” (Nickerson 2008: 394). The claim that any deviation from a normative principle of rationality can be interpreted either as an error performance or as due to a misunderstanding between the experimenters and the subjects is what Edward Stein (1996: 58-60) has called an “immunization strategy”, which defends the rationality of human beings from being subject to any empirical control. In other words, those who rely on such a strategy aim to argue that principles that are built in people’s reasoning competence correspond to those belonging to the *standard picture of rationality*, regardless of whether or not subjects’ answers to reasoning problems comply with what is dictated by those principles. However, what goes wrong with this argument is that the *standard picture of rationality* provides a narrow and highly constrained picture of human rationality as the normative principles of rationality derived from deductive logic, standard probability theory, and expected utility theory can be properly applied to people’s reasoning performances only in some restricted and very definite situations.
At this point, two main strategies can be used to develop an account of rationality alternative to the *standard picture*: (i) one can argue against the *standard picture of rationality* while assuming a deontological conception of rationality and, on the other hand, (ii) one can reject both the *standard picture* and the deontological conception replacing them with a consequentialist conception. These two strategies will be the subject-matter of the next chapters.
Chapter 2 – Non-standard deontological approaches to rationality assessment

Introduction
In the first chapter, I have characterized the Rationality and Irrationality theses in such a way that they both presuppose the standard picture of rationality. However, the issues related to human rationality are much more complex than the opposition between the Rationality and the Irrationality theses I have outlined there. On the one hand, it is true that many of the researchers involved in the rationality debate consider the standard picture of rationality as an absolute standard of correctness when evaluating human reasoning. On the other hand, many others have assumed and developed alternative accounts of rationality. After all, as pointed out in the previous chapter, the standard picture seems to provide a narrow and highly constrained picture of human rationality.

In this chapter, my focus is on two non-standard deontological accounts of rationality developed by Jonathan L. Cohen and Christopher Cherniak respectively. By the expression “non-standard”, I mean deontological accounts which do not assume the standard picture of rationality. Deontologism has been too often identified with such a picture, but in order to provide a comprehensive understanding of it, we have to look beyond that misleading identification. While the standard picture claims that to reason rationally is to reason in accordance with a set of normative principles derived from formal theories of reasoning such as deductive logic, standard probability theory, and expected utility theory, Cherniak and Cohen give opposed characterizations of rationality. According to Cohen, the normative standards of rationality against which to assess human reasoning should
ultimately be based on ordinary intuitions rather than on formal theories, whereas Cherniak, holding that it is impossible for humans to perform all the inferences the *standard picture of rationality* requires, redefines what it means to be rational on the basis of the limitations of the human mind. In the first part of the chapter, I provide a close examination of these two accounts by sketching out their main assumptions and purposes and then proceed to a critical assessment of them. In the second part, I examine the implications that the adoption of such accounts has for the evaluative question (for the characterization of the evaluative question see Chapter 1, Section 1.1). Finally, I suggest that any deontological account, either standard or non-standard, may be incomplete, or in other ways inadequate for assessing human reasoning.

### 1. Rationality and ordinary intuitions

In this section, I discuss a famous argument put forth by the philosopher Jonathan L. Cohen, according to which any attempt to characterize rationality should be based on people’s ordinary intuitions. The discussion falls into four main parts. First, I introduce the meta-normative account of rationality developed by Cohen, which is based on a process of mutual adjustment of normative principles to people’s ordinary intuitions, that is, a process of narrow *reflective equilibrium*. Second, I consider the explanatory role which Cohen attributes to people’s ordinary intuitions in characterizing the normative principles of rationality. Third, an application of Cohen’s method to empirical studies on probabilistic reasoning is presented. Finally, I examine the main weaknesses of this account and then proceed to consider two possible variants of it, that is, the wide and the expertise *reflective equilibrium* accounts.
1.1 Ordinary intuitions and reflective equilibrium

At the beginning of the eighties, the philosopher Jonathan L. Cohen argued against the alleged “bleak implications for human rationality” (Nisbett & Borgida 1975: 935) coming from the experimental findings on human reasoning (Cohen 1981; 1983; 1986: 149-192). His main thesis is that: “ordinary human reasoning – by which I mean the reasoning of adults who have not been systematically educated in any branch of logic or probability theory – cannot be held to be faultily programmed: it sets its own standards” (Cohen 1981: 317). In other words, he argues against the claim that human reasoning competence does not really fit with the normative standards of rationality. If human reasoning falls short of these standards, there is something wrong in them, not in human reasoning competence. But what does it mean that “human reasoning sets its own standards”? In virtue of what does certain reasoning count as rational? Cohen agrees with the supporters of the standard picture of rationality when they hold that rationality can be defined as the “validity in deductive or probabilistic reasoning” (Cohen 1981: 317). On the other hand, his proposal differs from classical accounts as to how to distinguish between valid and invalid reasoning. In order to define what counts as valid reasoning, logicians and probability theorists have traditionally referred to meta-logical theorems and mathematical proofs. But Cohen holds that the resulting theories do not provide a satisfying normative model of reasoning (Cohen 1981: 319). Against the received views, Cohen appeals to untutored people’s ordinary intuitions: any normative model of reasoning must be developed on the basis of this kind of intuitions.¹ More precisely, logicians should collect a large body of intuitions that ordinary people have about what counts as valid reasoning. These intuitions should then be abstracted and idealized in order to develop a normative model of rationality that prescribes how subjects ought to reason in order to count as rational. The resulting

¹ According to Cohen, an intuition that \( p \) is “just an immediate and untutored inclination, without evidence or inference, to judge that \( p \)” (Cohen 1981: 318). So the intuitions, which he refers to, amount to our non-reflective inclinations to judge that something follows from something else.
normative model will consist of the principles that best cohere with the idealized set of intuitions. According to Cohen, this method has a double target: it may lead logicians and probability theorists to revise the classical normative models of rationality, but it may also lead to modify our reasoning strategy by changing our ideas about what counts as valid reasoning in particular cases. In his view, these two targets can be met only through a process of mutual adjustment of normative principles to people’s ordinary intuitions, which should continue until what has been called a narrow reflective equilibrium is achieved. As such, this method (without calling it “reflective equilibrium”) was firstly developed by the philosopher Nelson Goodman (1955: 63-64).\(^2\) His goal was to devise a method for providing the principles of inductive reasoning with a justification. This process of justification has been charged with circularity by his critics, but Goodman always held that it is a virtuous circle because the principles of reasoning and our intuitions about what counts as good reasoning “are justified by being brought into agreement with each other” (Goodman 1955: 64).

Consider an example by Cohen (1981: 319). Suppose that we are faced with the following inference from (1) to (2) and asked to evaluate it:

\[
\begin{align*}
(1) & \quad \text{If John’s automobile is a Mini, John is poor, and if John’s automobile is a Rolls, John is rich.} \\
(2) & \quad \text{Either, if John’s automobile is a Mini, John is rich, or, if John’s automobile is a Rolls, John is poor.}
\end{align*}
\]

Intuitively, few people would accept this inference as valid. While most people would accept (1) as logically plausible, they, at the same time, would consider (2) as contradictory. However, according to the normative principles derived from deductive

\(^2\) It is noteworthy that Goodman’s method has been applied to justify not only the normative principles of inductive reasoning, but also principles involved in other philosophical disciplines. Most prominently, John Rawls (1971) has applied it in the field of moral theory. Rawls (1971: 21) himself has attributed the label reflective equilibrium to this method.
logic, the inference from (1) to (2) is valid. According to Cohen, in approaching this conflict between people’s ordinary intuitions and the classical normative model of reasoning, the method of narrow *reflective equilibrium* suggests that we should revise the latter or choose another normative model which complies with ordinary intuitions about this case. The new normative model should then be examined against further ordinary intuitions, and revised accordingly. In other words, whenever normative principles derived from deductive logic (and from the other classical formal theories of reasoning) differ from the ordinary intuitions of untutored people about what counts as valid reasoning, logicians should rely on the method of *reflective equilibrium* in order to assess and (if necessary) revise those principles. That can lead logicians to revise classical systems of logics, or even to devise different non-classical logics. So, according to Cohen, the process of *reflective equilibrium* iteratively optimizes the coherence between classical normative principles of reasoning and a given set of ordinary intuitions about what counts as reasoning validity, a process which can continue until a sufficient adjustment between principles and intuitions is reached.

### 1.2 Why do ordinary intuitions matter?

Why should it be advantageous to ground our account of rationality on ordinary intuitions? In order to answer this question, Cohen refers to the method developed by Noam Chomsky in linguistics (Chomsky 1965). The latter holds that, in developing and testing a theory in linguistics, linguists should rely on the ordinary intuitions that speakers of a language may have about the grammaticality of different kinds of linguistic structures: since our linguistic competence is not directly accessible, those intuitions are the main source of information about it. By relying on those intuitions, linguists should be able to identify the principles

---

3 Indeed, if people would accept that (1) implies (2), they would have to consider (2) to be as logically plausible as (1).
and rules of people’s linguistic competence. According to Cohen, just as a linguist starts with intuitions about the well-formedness of sentences and arrives at a description of people’s linguistic competence, so also researchers involved in the rationality debate aim to describe people’s reasoning competence. Accordingly, a descriptive account of human reasoning “[...] will describe a competence that normal human beings have – an ability, uniformly operative under ideal conditions and often under others, to form intuitive judgments about particular instances of right or wrong, deducibility or nondeducibility, probability or improbability” (Cohen 1981: 321). This part of Cohen’s analogy can be seen as an answer to the descriptive question that has been characterized in the previous chapter (see Chapter 1, Section 1.1). But that is not yet sufficient to explain the reason why we should ground normative models of rationality on ordinary intuitions. Cohen holds that this descriptive theory does double duty as a normative model. According to him, just as the grammar of a language can be used not only to describe but also to assess utterances in it, a theory of reasoning serves not only as a general descriptive theory about how people reason, but also as a normative model for assessing their individual ways of reasoning. Behind that, there is the idea according to which establishing what a rational individual ought to do, one should look first at what she actually does. At this point, there can be no gaps between normative principles of rationality and principles that are part of the human reasoning competence. Since both our normative principles of rationality and human reasoning are based on the same ordinary intuitions about what counts as good reasoning, Cohen argues that they cannot diverge; insofar as psychological experiments seem to prove that human reasoning systematically diverges from the accepted standards of rationality, they do so because either psychologists employ a wrong normative principle, i.e., a normative principle which is not in a narrow reflective equilibrium with people’s ordinary

4 In the sense that it can be taken to specify how the language faculty ought to operate in order to best carry out its function.
intuitions, or there have been interfering factors in the subjects’ reasoning performances. This last point recalls the distinction between competence and performance assumed by the supporters of the Rationality thesis that I have briefly presented in the previous chapter (see Chapter 1, Section 3.2). According to both the supporters of the Rationality thesis (committed to the standard picture of rationality) and Cohen, the supposed violations of the basic normative principles of rationality cannot be directly interpreted as indicating that those principles are not part of people’s reasoning competence. In their view, when a supposed violation occurs, there will be always an available explanation for keeping human reasoning competence safe: “[c]arelessness, conceit, wishful thinking, laziness, inattention, matching bias, and many other, less obvious factors are apt to make people’s inferential performance fall short of their inferential competence” (Cohen 1983: 511). Within such an explanatory framework, “nothing can count as an error of reasoning among our fellow adults unless even the author of the error would, under ideal conditions, agree that it is an error” (Cohen 1981: 322). So, according to Cohen, the conclusion to be drawn from the experimental data on human reasoning cannot be that humans are fundamentally irrational. However, he does not hold that people always reason correctly or that there is no error in their reasoning performances. Rather, he argues that empirical findings on human reasoning cannot show that people possess a faulty reasoning competence:

nothing in the existing literature on cognitive reasoning, nor in any possible future results of human experimental enquiry, could have bleak implications for human rationality, in the sense of implications that establish a faulty competence. At best experimenters in this area may hope to discover revealing patterns of illusions.

Clearly, the supporters of the standard picture of rationality and Cohen refer to two different sets of normative principles of rationality.

It is noteworthy that Cohen’s proposal is nowadays reflected the view of at least some analytic philosophers. For example, as is reported by Bishop and Trout, in discussing empirical researches on human reasoning, John Pollock and Joseph Cruz have recently stated that their position “is pretty much the same as the assessment of the irrationality literature offered by Jonathan Cohen” (Pollock & Cruz 1999: 130, n110; as reported by Bishop and Trout 2005: 127).
Often they will only be testing subjects’ intelligence or education. At worst they risk imputing fallacies where none exist. (Cohen 1981: 330)

In particular, Cohen has classified empirical studies that seem to demonstrate normatively inappropriate reasoning performances into four general cases:

- **Studies of cognitive illusions**: Cohen’s use of the term “cognitive illusion” differs significantly from that made by Tversky and Kahneman (Cohen 1981: 323-325; for Kahneman and Tversky’s use of the term see Chapter 1, Section 3.4). He argues that cognitive illusions should be attributed to some particular experimental conditions rather than to an alleged faulty reasoning competence on the part of the subjects. In particular, cognitive illusions would be the result of the use of (i) reasoning tasks that present problems which are very different from those people are faced with in their ordinary life. According to Cohen, the analogy with visual illusions amounts to deception. Cohen holds that experimenters usually change the features of the experimental setting “[...] in such a way that subjects are induced to indulge in a form of reasoning that on a few moments’ prompted reflection they would be willing to admit is invalid” (Cohen 1981: 323).

- **Tests of intelligence or education**: Cohen argues that some violations of norms of rationality are due to limited education, not to irrationality. In other terms, some of the experimental findings on human reasoning should be considered as indicators of lack of formal knowledge in logic and probability theory rather than of systematic irrational behaviour. On this view, what is needed in order to improve our reasoning strategies is a formal education. So, experimental findings on human reasoning would have only pedagogical implications (Cohen 1981: 325-326).

- **Misapplications of an appropriate normative model**: although the experimenter may be assuming a justified normative model of rationality, the subject might be
interpreting the reasoning problem she is presented with in a way different from the experimenter’s, and therefore be providing a correct response (according to the model assumed by the experimenter) to a different problem (Cohen 1981: 326-328; see also Stanovich 1999). In such cases the experimenter does not seem to recognize that there were information available in the experimental context that might lead subjects to interpret the reasoning problem in a way that differs from what she assumed to be the case;

- *applications of inappropriate normative models*: in contrast to the previous case, this case includes situations in which it is assumed that the subject has interpreted the task as the experimenter assumed to be the case, but the latter evaluates the subject’s response against an inappropriate normative standard (Cohen 1981: 328-330). This kind of cases will be discussed in the next section (see Chapter 2, Section 1.3).

We can divide these four cases into two main categories. In the first two cases, although the subjects’ responses have been shown to be normatively inappropriate, the empirical data gathered by the experimenters are deemed not to reveal anything about their reasoning competence. In the second two cases, subjects’ responses are apparently wrong and the error is attributed to the experimenter. According to Cohen, however, in any of those cases no one is allowed to conclude that human reasoning competence is normatively problematic.

1.3 Ordinary intuitions and probabilistic reasoning

Cohen takes intuition-based rationality as fundamental when evaluating human reasoning. His proposal appears to be in line with received attempts to develop logical systems and probability theories. Indeed, these theories have been traditionally aimed at systematize
people’s intuitions about what counts as good reasoning. As experimental researches on human reasoning show, however, the resulting theories go beyond and even conflict with people’s intuitions and their inferential practices. So, according to Cohen, the classical formal theories of reasoning do not appropriately systematize people’s ordinary intuitions and consequently the standard picture of rationality should be rejected, rather than considering ordinary intuitions about what counts as valid reasoning as incoherent. Consider the case of probabilistic reasoning (Cohen 1979; 1981: 328-330; 1986: 157-192). In Kahneman and Tversky’s works, it is presupposed that a given reasoning problem has a single correct answer, based on the normative principles derived from the Bayesian (or subjective) probability theory. As seen in the previous chapter (see Chapter 1, Section 2.1), these normative principles are justified a priori by relying on seemingly indisputable arguments. Cohen holds that none of them establishes which conception of probability is operative in ordinary people’s reasoning performances:

in order to discover what criteria of probability are appropriate for the evaluation of lay reasoning we have to investigate what judgments of probability are intuitively acceptable to lay adults and what rational constraints these judgments are supposed to place on one another. (Cohen 1981: 319)

In other words, according to him, normative models cannot be selected a priori, but should be developed starting from the intuitions and reactions of ordinary people, such as those of the participants in the experiments. Look now at the experimental research devised by Kahneman and Tversky. In their reasoning tasks, participants were asked to attribute probabilities to individual events or outcomes. As pointed out above, Kahneman and Tversky assume the normative principles derived from the Bayesian (or subjective) probability theory as the only appropriate normative standard for evaluation. With regard to
the probabilities of single outcomes or events, Cohen argues that there are not only one but rather two kinds of normative models, one based on the Pascalian conception of probability (which he identifies with the Bayesian probability theory) and the other following from the Baconian conception. Whereas, according to the Pascalian conception, probabilities should be attributed “on the assumption that all relevant facts are specified in the evidence”, the Baconian conception is described by Cohen as grading probabilities “by the extent to which all relevant facts are specified in the evidence” (Cohen 1979: 389; as reported by Todorov 1997: 401). Let me explain the difference between these two models. The use of a Pascalian conception of probability is appropriate only when one estimates the probability that an event will occur or has occurred on the basis of a collection of similar events. Such a judgment can be made on the basis of the frequency one has experienced certain events in the past or on the basis of an a priori assessment, such in the case of lotteries. Looking at situations that involve decision-making, however, you can see that such situations are usually unique and so the Pascalian conception of probability cannot be properly applied to them. Cohen holds that in trials and unique decision-making situations people’s reasoning performances can be explained and evaluated only on the basis of the Baconian conception of probability. Drawing a general conclusion, Cohen holds that, when evaluating one’s probabilistic judgments, experimenters should assume the appropriate conception of probability. In particular, when the reasoner has been frequently faced with a given situation before, the Pascalian conception should be assumed, whereas the Baconian is appropriate when the situation is unique or it is the first time the reasoner meets it. In reasoning experiments, classically, subjects are asked to attribute probabilities to outcomes or events which they meet for the first time (or are unique according to them). If so, the Baconian conception of probability comes closer to how subjects reason in the experimental situation than the Pascalian one. Cohen concludes that subjects’ “irrational”
responses to reasoning tasks fit nicely with a Baconian interpretation (Cohen 1981: 330). For example, he argues that the alleged *representativeness heuristic* (as characterized by Kahneman and Tversky)

leads to a conclusion about probability in a sense in which an inference from representativeness to probability is always quite legitimate – albeit a sense that conforms to principles different from those derivable within the calculus of chance [i.e., principles that are derivable from the Baconian conception of probability]. (Cohen 1981: 330)

By assuming an appropriate conception of probability, it is possible to demonstrate that, in certain situations, probabilistic judgments based on representativeness are normatively appropriate. By way of conclusion, Cohen maintains that, since the appropriate norms of rationality are already immanent in people’s inferential practices, in order to argue for human rationality a supporter of the Rationality thesis should change her mind and assume a Baconian conception of probability.

1.3 Ordinary intuitions and relativism

Needless to say, Cohen’s proposal has been and continues today to be widely discussed and criticized.⁷ I would like to focus here on one of the main questionable consequences of it (see also Nickerson 2007: 405). One of his main opponents, Daniel Kahneman, has pointed out that Cohen’s attempt to characterize rationality on the basis of ordinary intuitions is not in principle completely wrong: he admits that “one of the criteria for a norm of thought and action must be that reasonable people will want to obey it”. From that, it follows that “reasonable people must be able to recognize a rational argument, and in this restricted

sense that they must be rational” (Kahneman 1981: 340). However, he also claims that “it is improper to argue, as Cohen does, from this general belief in human rationality to a belief in the rationality of any notion for which a majority can be found” (Kahneman 1981: 340). Similarly, Henry Kyburg (1981: 342) has observed that, when determining what rationality requires, we ought not “to honour all intuitions indiscriminately – even in qualified form – but to reduce collections of intuitions from which others can be derived”. Indeed, ordinary people, as well as logicians, do not usually agree on what counts as rational in particular cases, and a given person may change her opinion with time. Many people have experienced to change their minds about the validity of certain instances of reasoning after having reflected on them. What empirical findings on human reasoning show is that different subjects usually give different responses in the same reasoning task. That means that they have different intuitions about what is normatively appropriate in that task. For example, looking at the experimental results of the standard version of the Wason selection task, one can find at least four different kinds of responses offered by the subjects. But which of these is grounded on the right intuitions about what is normatively appropriate? Who can tell us which intuitions are the best? How should we in the end reach a reflective equilibrium between these different sets of intuitions and the supposed normative principles? Researchers who consider ordinary intuitions as the appropriate basis for defining rationality usually deny both that people’s ordinary intuitions may really differ from one another and that people may really reason badly from their premises. 8 In particular, Cohen points out that normal adult people’s reasoning competences cannot differ among them. On the one side, he accepts that “two different people, or the same people on two different occasions, may sometimes have apparently conflicting intuitions” (Cohen 1981: 319). On the other side, however, Cohen holds that these differences are only

8 Cohen (1983: 512) claims that “[...] basic normative theory has to be grounded on intuitions that are supposed to be universally available”.
apparent and so there are no real conflicts (these apparent conflicts can be easily reconciled). For example, he explains,

> the people involved might come to recognise some tacit misunderstanding about
> the terms of the problem, so that there is no real conflict; or they might repudiate a
> previously robust intuition, perhaps as a result of becoming aware that an otherwise
> preferred solution has unacceptable implications [...]. (Cohen 1981: 319)

Many experimental results seem to demonstrate the opposite, that is, subjects usually stick to their intuitions and beliefs even after receiving several suggestions and advice by the experimenter. It seems to me that only insofar as intuitions about reasoning validity are perfectly uniform across different people, Cohen’s method could provide a set of stable and firm normative standards against which human reasoning can be assessed. Otherwise, deciding which normative standards to adopt when evaluating reasoning performances becomes an impossible task. Indeed, without uniformity among ordinary intuitions, there is no possibility to develop stable and firm normative standards, thereby leading to relativism in rationality assessments.

Moreover, if we have assumed that people’s ordinary intuitions about what counts as good reasoning are uniform, it might be interesting to consider what kinds of principles are in *reflective equilibrium* with those intuitions. As Stephen Stich (1990: 86) remarks, when strictly applied, the method of *reflective equilibrium* may lead to identify plenty of clearly unacceptable normative principles. For example, when playing games of chance, people usually reason in accordance with what is known as the gambler’s fallacy. That is, they judge that “the likelihood of throwing a seven in a game of craps increases each time a non-seven is thrown” (Stich 1990: 86). In such cases the principle underlying their judgments is in *reflective equilibrium* with their ordinary intuitions. Indeed, if such a principle is articulated and people consider it and the judgement which they have made in accordance
with it, they will probably accept both. But the general normative principle we may derive from their intuitions is based on the gambler’s fallacy and so, in conforming to it, they will make an irrational judgment. Nonetheless, if we assume Cohen’s account of how principles of rationality should be derived, these people will not be regarded as irrational. Since the resulting principle endorsing the gambler’s fallacy is not rational, Stich (1990: 86) concludes that Cohen’s method is shown to be flawed. Indeed, if a given reasoning is distinctly irrational or counter-intuitive, and if it accords with a subject’s reasoning competence, then we should conclude that her reasoning competence is normatively problematic.

1.4 Wide reflective equilibrium and experts’ intuitions

As seen above, according to Cohen (1981: 320), human reasoning must be evaluated in its own terms and by its own standards. To the extent to which the derivation of the normative principles of rationality is determined by appealing only to ordinary intuitions, Cohen’s method is characterized as narrow reflective equilibrium. It is called “narrow” because there is no other element that comes into the process of adjustment between normative principles and ordinary intuitions. However, among those who maintain that the norms of rationality should be derived by a process of reflective equilibrium, there are some who disagree with Cohen on the idea that this process must be a narrow one. On the one side, scholars such as Edward Stein (1996) have proposed that the process of adjustment should involve not only normative principles and ordinary intuitions, but also epistemological, metaphysical, and neuropsychological theories. This kind of reflective equilibrium has been called wide reflective equilibrium. On the other side, other scholars have maintained that the process of reflective equilibrium should involve a very restricted set of intuitions, that is, only the intuitions of the experts in the relevant disciplines (Stich & Nisbett 1980). Finally, Paul
Thagard and Richard Nisbett (1983) have proposed a mixed method of reflective equilibrium that takes into account both general theories and experts’ intuitions.

Firstly, according to Edward Stein, a wide reflective equilibrium may be achieved when a set of ordinary intuitions, a set of normative principles, and a set of general theories, such as neurophysiologic and epistemological one, are brought into agreement. This method leads to what Stein (1996: 255) calls the naturalized picture of rationality. As he points out, in justifying normative principles of rationality, we should always take into account “scientific evidence about humans that bears on which principles of reasoning are computationally accessible to us” (Stein 1996: 255). Stein’s proposal, however, seems to be questionable. Indeed, before any empirical evidence is used in attaining a set of normative standards of reasoning, its relevance to the normative question of rationality should be assessed. As Engel (1998: 388) points out, “this evidence does not come norm-free, and we must have some standards about reasoning to assess whether given data are relevant to reasoning”. But if scientific evidence comes into the process of reflective equilibrium, it is not clear what standards or principles should be previously applied to determine whether or not that scientific evidence is relevant to the principle we are trying to justify.

Secondly, according to the expert version, a principle counts as normative “if it is in reflective equilibrium for those people in a position to assess the relevant considerations” (Stein 1996: 147; see also Nisbett & Stich 1983: 198-200). As a consequence, a principle is justified when it comes from a process of reflective equilibrium based on experts’, not ordinary intuitions. In order to determine normative principles of rationality, we have to appeal to intuitions about what counts as good reasoning for a definite group of persons, namely, the ones who are considered as having a certain authority in the field of research in question.⁹ In favour of this kind of reflective equilibrium, it is argued that it prevents that

⁹ Cohen (1981: 320) would not agree with this view as “the judgments of everyday reasoning must be evaluated in their own terms and by their own standards”, not by experts’ standards.
“counter-intuitive” principles, such as the gambler’s fallacy, may count as normative. As is summarized by Raymond Nickerson (2007: 42), two main objections can be levelled against this suggested modification of the process of *reflective equilibrium*. First of all, this modification raises the question of how to determine what the relevant disciplines are to which the experts should belong. Secondly, with regard to the rationality debate, there is the fact that scholars trained in the relevant disciplines, such as logicians and epistemologists, do not usually agree on what the normative principles of their disciplines are. For example, the commentary on Cohen’s target article (1981) provides us with compelling evidence of such disagreements. As is pointed out by Nickerson (2007: 42), in all the commentary what an expert in a particular discipline takes as an example of rational reasoning is seen by another as an example of normatively inappropriate reasoning.

2. Minimal Rationality

Instead of appealing to people’s ordinary intuitions, those who want to move away from the *standard picture of rationality* might refer to people’s cognitive limitations. Specifically, an interesting analysis of how human rationality should be conceived, given the cognitive limitations of the human mind, has been developed by the philosopher Christopher Cherniak in the mid-eighties. According to him, the kind of rationality that is applicable to creatures with limited cognitive resources and time, as humans are, is a minimal rationality,\(^{10}\) which Cherniak distinguishes from complete lack of rationality and the possession of an ideal rationality.

\(^{10}\) A recent account of minimal rationality, different from that developed by Cherniak, has been proposed by Fred Dretske (2006).
2.1 The human finitary predicament

According to Cherniak (1986: 8), the most fundamental aspect of the human condition is that “humans are in the finitary predicament of having fixed limits on their cognitive capacities and the time available to them”.\(^{11}\) Humans have very limited time, computational capacities and working memory. Because of these limitations, they cannot perform all the inferences the classical normative principles of rationality require. Insofar as the inferences they require to do are not feasible, it is not so obvious that the classical normative models of rationality can really be considered “normative” for human beings. This criticism is in line with what is well-known in ethics as the “ought-implies-can” principle, according to which normative principles, as such, cannot prescribe to do what we cannot actually do (Cherniak 1986: 113). Obviously, the relevance of a normative principle depends on what is taken to be within one’s cognitive capacity. In his examples, Cherniak usually refers to issues of checking one’s beliefs for consistency. According to him, a complete check of one’s current set of beliefs for truth functional consistency would lead to a computational explosion.\(^{12}\) As usually happens, indeed, when an individual tries to solve contradictions among her beliefs, that does not actually involve a process of exhaustive search. People are faced with temporary inconsistency continuously, whenever something unexpected occurs, and that does not seem to be a limitation of their rationality (see, e.g., Harman 1986; 1995: 18-21).

2.2 To be minimally rational

Which consequences does the human finitary predicament have in philosophical theorizing about human rationality? Cherniak (1986: 9) begins by examining two opposite

---

\(^{11}\) Cherniak puts forth both empirical observations and theoretical analysis in explaining this point.

\(^{12}\) Cherniak (1986: 93-94) has calculated that, in order to make a complete check for truth functional consistency a belief system containing 138 logically independent propositions would require the power of a super computer which should work (in order to complete this task) for more than twenty billions years.
conceptions of rationality, based on what he calls respectively null and ideal rationality conditions. He rejected both on the grounds that the null rationality condition does not yield any standard against which to predict another’s behaviour (and more generally, any benchmark against which to assess human reasoning), while the ideal rationality condition poses too high standards for beings, as humans are, with limited cognitive resources and time.\textsuperscript{13} Setting aside these two conceptions of rationality, Cherniak proposes what he calls minimal rationality, according to which a rational subject should be able to conform to at least some of the classical normative principles of rationality. More generally, in order to count as rational people should solve at least some of the inconsistencies in their belief sets, they should perform at least some feasible inferences that are relevant for them, and the like. It would be an error to think that Cherniak’s theory does not provide any criteria that can be employed to distinguish between better and worse ways of reasoning. Rather it assumes that rationality assessments are not an all-or-nothing affair. As Cherniak (1986: 24) puts it, “to be minimally normatively rational, we must take into account not only (1) the soundness of the inference, but also (2) its feasibility and (3) its apparent usefulness according to the agent’s belief-desire set”.\textsuperscript{14} In other terms, according to him, given our cognitive limitations, “not making the vast majority of sound and feasible inferences is not irrational; is rational” (Cherniak 1986: 24). As is pointed out also by Harman (1995: 22), it is not always a rational move to react to the finding of an inconsistency by abandoning every other activity in order to engage in eliminating that inconsistency.

Much of what seemed to be irrationality in reasoning experiments can be shown as instances of rationality on the basis of minimal rationality. Coming back to the heuristics

\textsuperscript{13} In the previous chapter, I have introduced three accounts of behaviour interpretation and propositional attitude attribution and discussed the role that assumptions about rationality plays in them (see Chapter 1, Section 3.5).

\textsuperscript{14} Moreover, according to Cherniak (1986: 31), “it is false that the difficulty of any given inference is always the same for all reasoners, or even all normal humans. Consequently, the difficulty of an inference cannot correspond to any syntactical feature that is intrinsic to the inference, in isolation from the reasoning psychology of the deducer”.

69
(see also Chapter 1, Section 4.4), the supporters of the Irrationality thesis argue that insofar as heuristics do not lead reliably to rational behaviour, it is not possible to count them as rational processes. From the point of view of minimal rationality, the fact that people rely on heuristics should not be considered as a demonstration of people’s irrationality but rather as a good attempt to cope with their “finitary predicament”. While relying on heuristics has certainly positive effects for people’s cognitive economy, people should be conscious that those heuristics lead to reasonable but uncertain solutions. So, minimal rationality requires of people that they ought not always to reason in accordance with the classical principles of rationality, not only because, as we have seen, they cannot do that, but also because certain attempts to conform to those principles would make them waste too much time and energy. In this sense, rationality should be understood as the more efficient way of outweighing the benefits that could be gained from using a reasoning strategy compared to the costs of using it. That consideration opens the question of how to weight benefits and costs in assessing reasoning strategies. This point will be tackled in the next section (see Chapter 2, Section 2.3).

To sum up, Cherniak (1986: 86) maintains that his minimal conception of rationality permits us to acknowledge the basic platitude that human beings are in the finitary predicament, and therefore ought to use some such heuristics — [...] formally incorrect heuristics need not in fact be irrational at all.

According to minimal rationality, which kinds of inference we can expect a subject to make is mainly a function of that subject’s cognitive resource allocation.

2.3 Relevance and context

A fundamental role in Cherniak’ account, although not explicitly stated, is played by the notion of relevance. Following his analysis, we can hold that, according to minimal
rationality, people count as rational insofar as they derive those inferences that are most relevant for them. In other words, to be rational does not mean that a rational subject always ought to draw all the normatively appropriate inferences, rather, she ought to draw those among them that are more relevant for her. As Harman (1986; 1995) has argued, neither deductive logic nor standard probability theory specify what kinds of inference are relevant in a given circumstance and what the relevant premises of our reasoning are. In particular, normative principles derived from deductive logic specify what kinds of conclusion can be validly inferred from a given set of premises, but they do not suggest what a reasoner ought to believe, given a set of new information. As we have seen in the previous chapter (see Chapter 1, Section 5), this is one of the main reasons why the standard picture of rationality provides a narrow and highly constrained picture of human rationality.

The question of relevance is strictly related to the context-dependence of assessments of reasoning performances. According to Clifford Hooker (1994), context-dependence is a crucial aspect in Cherniak’s account, although he has never developed this aspect deeply enough. As seen before, Cherniak holds that it is rational for people to rely on heuristics which he considers as cognitive shortcuts that are very useful for reaching quick even if uncertain solutions. Indeed, where the costs of employing a reasoning strategy based on principles of logic or probability theory outbalance the benefits that could be gained from using it, rational subjects should rely on sub-optimal reasoning strategies and, according to Cherniak, that is the most rational thing to do. On the one hand, the question of resource allocation, such as benefits and costs in using a particular reasoning strategy, plays a fundamental role in assessing human reasoning. On the other hand, as Hooker (1994: 205-206) has pointed out, the choice of a specific strategy is also context-dependent because the cognitive resources available to subjects, the reasoning strategies relevant for the problem
they are faced with, and the benefits and costs involved in using it all typically vary with context. Accordingly, costs and benefits are context-dependent and that should lead to the context-dependence of human reasoning. Consider an example by Cherniak. Contrary to what the standard picture of rationality holds, he argues that the normative principle of rationality concerning the modus ponens should not always be applied. Its application depends on the circumstances. According to Cherniak, only in situations in which a subject believes both that \( p \) and that \( p \) implies \( q \) and is interested in whether \( q \) holds, she ought to apply the normative principle concerning the modus ponens. These are situations where the inference required to be made is feasible to the reasoner and is “positively useful [for her] at a given time” (Cherniak 1986: 24). Many other inferences are inferences that are normatively appropriate but not reasonable (not useful) for a subject to make in a certain context at a given time. If we draw all the conclusions from Cherniak’s argument, then we should assume that the norms of rationality become not only psychologically dependent, as Cherniak correctly points out, but also robustly context-dependent.

3. Non-standard deontological conceptions and rationality assessment

This section is organized in two parts. In the first part, I consider the implications that the adoption of a non-standard deontological account of rationality has for what we have called the evaluative question, by examining the rationality conditions for individuals that such accounts assume. I argue then that both intuition-based and minimal accounts of rationality provide a very partial and incomplete response to the evaluative question. In the second part, I examine the idea of a “prescriptive” model of rationality, which is characterized as a resource-relative deontological account of rationality, and then criticize it. Finally, I raise a more general objection against any deontological account, either standard or non-standard.
3.1 The rationality condition for individuals

What implications does the adoption of such views have for the evaluative question? In the previous chapter, I have maintained that every deontological account of rationality assumes implicitly a definite general requirement, which connects an individual’s capacity to reason, implicitly or explicitly, in accordance with the relevant norms with her rationality (see Chapter 1, Section 3.3). And I have added that very different characterizations of this requirement may be developed: it may state either necessary and sufficient conditions or be weakened to some degree. In particular, the requirement implicitly assumed by the supporters of the standard picture of rationality is the following:

(R1) Given a definite set $N$ of norms of rationality, a subject $S$ qualifies as rational if and only if $S$’s behaviour conforms to $N$.

This requirement articulates a necessary and sufficient condition in order for $S$ to qualify as rational. To fulfil it, $S$ never has to depart from what is dictated by the norms belonging to $N$: according to it, rationality is taken as an all-or-nothing affair. Consider now the account of rationality developed by Jonathan Cohen. What he offers is a method based on a process of mutual adjustment of normative principles to people’s ordinary intuitions. In other words, Cohen tries to entirely answer the normative question by providing both (a) a set of normative principles of reasoning and (ii) a meta-normative account of how these principles can be justified and (if necessary) revised.\footnote{In the previous chapter, I have held that the normative question calls for a general normative model of how subjects ought to reason in order to qualify as rational (see Chapter 1, Section 1.1).} Characterizing the normative principles of rationality through this method, Cohen concludes that these principles correspond to those that are part of people’s reasoning competence. So, provided both (i) his deontological conception of rationality and (ii) that he assumes that it is impossible that people are irrational (the normative principles of rationality correspond to those that are part of human reasoning competence), it could be argued that Cohen, as well as those who are committed
to the *standard picture of rationality*, assumes (R1). In particular, in assuming (R1), the supporters of the Irrationality thesis and Cohen draw different conclusions because of the different content which they assign to $N$: on the one hand, having assumed the *standard picture of rationality*, the supporters of the Irrationality thesis claim that people are irrational, that is, they do not possess a normatively appropriate reasoning competence, and, on the other hand, Cohen, assuming only the norms of rationality which are sanctioned by people’s ordinary intuitions, argues that people are always rational.

Cherniak provides a completely different argument to defend human rationality. According to him, it is not sufficient to lay down a general requirement for rationality as long as people cannot satisfy it. Hence, Cherniak’s response to the consequences of the human finitary predicament is weakening the general requirement for rationality without changing its formal character. Indeed, his proposal can be rephrased by replacing the “if and only if” in (R1) with a simple conditional form and adding some specifications:

/(R2) Given a definite set $N$ of norms of rationality, a subject $S$ qualifies as rational if $S$’s behaviour conforms to at least some norms belonging to $N$.

According to (R2), people need not to conform to all the norms of rationality because of their cognitive limitations. Provided that in many contexts the cost of conforming to the classical normative principles of rationality outweighs any benefits gained from doing so, (R1) seems to be too strong a requirement on human reasoners. That is why people ought to conform only to some norms of rationality, i.e., those which are apparently relevant to the task of inferring the more appropriate conclusions, or making the more appropriate decisions, etc. In other words, every human should conform to at least some norms of rationality according to her actual cognitive situation, her beliefs, her desires etc. But reasoning in accordance with the norms belonging to $N$ is not always the more rational thing to do: it does not, in itself, guarantee rationality. A rational subject can conform to
these norms selectively, that is, she can conform to them when it is useful for her goals and, if it is not, she can ignore their normative force. So, to be minimally rational is to be more disposed to do what is rational, relative to one’s cognitive situation, than what is not.

### 3.2 What goes wrong with intuition-based and minimal rationality?

In my view, both intuition-based and minimal accounts of rationality seem to provide a very partial and incomplete response to what I have called the evaluative question, that is, the question of how to select the appropriate normative standard when assessing a given reasoning performance and how to apply it appropriately.

Beginning with Cohen’s answer to the evaluative question, I would like to distinguish between two different interpretations of his intuition-based account of rationality: while, according to the first interpretation, rationality amounts to the ability to conform to people’s prevailing intuitions about what counts as reasoning validity, the second one characterizes rationality as relative and calls for a comparative evaluation among different normative models based on competing ordinary intuitions about reasoning validity. Consider the first interpretation. If we accept the first interpretation combined with an optimistic interpretation of reasoning performances, assessing people’s reasoning performances becomes pointless. Indeed, insofar as both the norms of rationality and human reasoning are based on the same ordinary intuitions about what counts as valid reasoning, there seems to be no difference between how people reason and how they should reason. If so, assessing human reasoning appears to be dependent on people’s prevailing intuitions about reasoning validity. There is a danger here that the notion of rationality is deprived of all of its normative force: if being rational amounts to one’s conformity with other’s prevailing intuitions, in order to count as rational all people must reason in the same way, regardless of whether or not this really is the best way to reason (see also Oaksford & Chater 2007: 75)
24). But look at the second interpretation of the intuition-based account of rationality. As is pointed out by Samuels and his colleagues (2004: 140), the analogy with linguistics proposed by Cohen is deceptive, after all. Linguists do not aim at assessing people’s linguistic competence. Indeed, speaking different languages, people belonging to different culture and societies have different linguistic competences and such differences are made explicit in their linguistic intuitions. People who speak Italian as their first language have linguistic competences that are very different from that of an Englishman, and her linguistic competence in turn differ from that of, for example, a Greek speaker. The question of which among these kinds of competence is best or which one is the correct one does not seem to make sense. Richard Samuels and his colleagues (2004: 140) observe that

on the language side of the analogy, there are performance errors, but there is no such thing as a competence error or a normatively problematic competence. If two otherwise normal people have different linguistic competences, then they simply speak different languages or different dialects. On the reasoning side of the analogy, however, things look very different. It is not clear whether there are significant individual and group differences in the rules and principles underlying people’s performance on reasoning tasks, as there so clearly are in the rules and principles underlying people’s linguistic performance.

If in Cohen’s account a normative model of reasoning corresponds to a descriptive account of reasoning competence (they are both developed on the basis of the same intuitions and idealized in the same way), and if people have very different intuitions about reasoning validity and different reasoning competences, then there are many normative models of reasoning and it appears to make sense to ask which normative model is the right one (see also Stich 1985: 131). But what are the criteria to evaluate normative models? Ordinary intuitions cannot be, so what? As seen above, according to Cohen, ordinary intuitions are
non-reflective inclinations to judge that something follows from something else. In contrast, it seems to me that ordinary intuitions about reasoning validity are always made against a background of previous knowledge and are influenced by both the type of questions and the type of circumstances. As is shown by empirical studies on human reasoning, context and content of a given reasoning task strongly influence the subject’s conclusions. If so, an interesting line of enquiry may be to find which contextual factors support the use, for example, of a particular conception of probability (either the Pascalian or Baconian one) and consequently of the use of a given normative model in a given situation. So, in order to give an answer to the evaluative question, a supporter of this view should specify the features of a situation which support the use of a particular normative model of rationality. While Cohen gives no answers to this question, Cherniak tries to reply to it by providing a more complex account about how to select the normative standards. However, his minimal rationality does not reach the target either. Looking at the contrast between his minimal rationality and the standard picture of rationality, Cherniak assumes that humans need in the first case to meet the minimal normative standards of rationality, while in the second one to meet the ideal standards of rationality. So, in Cherniak’s view, irrationality means non-rationality in the sense that those who are charged with irrationality are regarded as not having met the basic conditions necessary for rationality (see Hanna 2006: VII). But which are the basic conditions necessary for rationality? How could Cherniak’s proposal account for those conditions? Putting the question in other words, according to his account, the main question is what sound and feasible inferences a subject ought to draw in order to count as rational. While suggesting that what a rational subject ought to do depends on her actual cognitive situation, Cherniak does not offer an account of how to determine the basic conditions necessary for rationality. For example, he makes no

16 In the same vein, Jonathan Weinberg, Shaun Nichols e Stephen Stich (2001) have empirically demonstrated that people’s intuitions about what counts as belief or knowledge strongly diverge and these divergences are due to both culture and the socioeconomic status of the people providing the intuitions.
attempt to characterize the notion of “at least some” more closely (see above the rationality condition (R2)). He admits that his rationality condition is too vague but points out that vagueness has some advantages (Cherniak 1986). According to Cherniak, it is an open question whether or not there are particular inferences (even the most “obvious” ones, like the *modus ponens*) which should be included in the minimal condition. As we will see later (see 3.4), however, because of the vagueness of the minimal rationality condition, a supporter of minimal rationality is faced with several problems in the attempt to develop a satisfactory deontological normative framework for assessing human reasoning.

### 3.3 From normative to prescriptive models of rationality

An alternative solution for non-standard deontological accounts is to weaken the classical normative principles of rationality. The *standard picture of rationality* seems to be inapplicable to account for the real reasoning performance of people, so why should experimenters try to impose it in assessing subjects’ performances on reasoning tasks? As seen above, failures in human reasoning should not always be considered as evidence of people’s irrationality. When assessing human reasoning, constraints on human ability to reason have to be taken into account. As is pointed out by Gilbert Harman (1995: 12), reasoning

uses resources and there are limits to the available resources. Reasoners have limited attention spans, limited memories, and limited time. Ideal rationality is now always possible for limited beings. Because of our limits, we make use of strategies and heuristics, rules of thumb that work or seem to work most of the time but not always. It is rational for us to use such rules, if we have nothing better that will give us reasonable answers in the light of our limited resources.
These considerations should lead us to identify realistic norms for human subjects, that is, norms computable by their cognitive systems. The attempts to model the principles of rationality to the limited resources of the human cognitive system are characterized as prescriptive models of rationality (see, e.g., Baron 2008). As Keith Stanovich (1999: 4) observes, two main cases can be distinguished. In the first case, when a given normative principle is computable by the human cognitive system, it can be also regarded as prescriptive. In the second case, when the normative principle is not computable, the appropriate standard for human reasoning performance is the computable strategy closest to that normative principle.

It is noteworthy that those who develop prescriptive models of rationality do not refuse the classical normative model of rationality. According to Jonathan Baron (2008: 32-34), normative models can be employed as appropriate standards against which to assess prescriptive models. On this view, to reason in accordance with a prescriptive model is enough to count as rational but a good prescriptive model should not depart too much from an appropriate normative model. Accordingly, a prescriptive principle can be characterized as a principle which can help us to get closer to the appropriate normative model. As is exemplified by David Over (2004: 12),

there may well be an ideally healthy diet for each of us given our individual genetic make-ups, but it would be almost impossible for us to find out what this ideal is and follow its rules. We may, however, be able to follow prescriptive “rules of thumb”, like trying to cut down on foods high in fat and sugar.

While it is trivially easy to conform to certain classical normative principles of rationality, the standard picture of rationality itself can only be regarded as an ideal model of
rationality. Prescriptive principles help people approach the *standard picture* to some extent by showing when to rely on a reliable heuristic strategy rather than on its principles.

Developing a prescriptive model for human reasoning is not so easy a task as its proponents appear to hold. Starting from a normative model, prescriptive models can be developed with very different purposes, all depending on what kinds of constraints are assumed to operate when people reason. In particular, there are at least three different kinds of constraints which can be identified as relevant to developing prescriptive models for human reasoning: (i) *a priori*, abstract constraints, which are due to the computational properties required to solve certain reasoning tasks. In particular, there are reasoning tasks (e.g., the traveling salesman problem) that are *a priori* computationally intractable and thereby impossible to solve, regardless of how our cognitive systems work; (ii) cognitive constraints, which reflect the fallible and limited nature of our cognitive system. People cannot solve certain reasoning tasks because of their not having efficient cognitive algorithms to handle them. These constraints are universal across humans; (iii) idiosyncratic, subjective constraints that depend on people’s individual cognitive capacities, their reasoning abilities, their actual cognitive situations, etc.\(^{17}\)

When dealing with prescriptive models of rationality, all three kinds of constraints should be taken into account. While, however, (i) and (ii) are to a certain extent objective, (iii) seems to call for defining different standards for different reasoners with different cognitive capacities and reasoning abilities, but it is difficult to see then how to manage that. So, the key question here is whether one or more than one prescriptive model is required to be developed in order to evaluate reasoning performed by people with different cognitive capacities and reasoning abilities.

\(^{17}\) Looking at human beings, one can see that they are very different one from another both in their reasoning abilities and cognitive capacities; indeed, children usually are more cognitively limited than adults, and reasoning abilities and cognitive capacities of people with an high IQ differ from those whose IQ is lower (see Stanovich 1999).
3.4 The blind spot of any deontological account

Any deontological account of rationality, either standard or non-standard, faces with a more general problem. While, according to deontologists, good reasoning should be characterized in terms of its conformity to a specific set of normative principles $N$ (which are regarded as the normative principles of rationality), they do not explain why people must conform to those principles, namely why people should be interested in reasoning rationally (Samuels et al. 2004: 166). Some deontologists seem to assume that there could be evolutionary or psychological reasons which explain why people ought to conform to them. However, they do not offer any compelling argument in support of their claims. Other deontologists, instead, hold that the reason why people must conform to the principles belonging to $N$ is that only conforming to them will reliably lead to good outcomes or, as Jonathan Baron (2008: 319-320) explains, following them is in the long run the most successful way of achieving one’s goals. This seems to be a more compelling reason. As is argued by Samuels and his colleagues (2004: 166), however, by stating that to reason in accordance with some set of normative principles $N$ is what characterizes human rationality, deontologists assumes that a subject $S$ who always reasons in accordance with $N$ is reasoning rationally and so is rational regardless of whether there are more efficient ways to achieve the goal at which she aims. As I understand their claim, according to deontologists there are good reasons to follow the normative principles characterizing rationality, regardless of whether there are better and feasible ways to reason. Accordingly, if there are situations in which reasoning in accordance with $N$ is not the most efficient way of achieving a particular goal, it would be irrational to use a more efficient reasoning strategy, which falls short of what is stated by the alleged relevant principle of rationality, for achieving that goal (Samuels et al. 2004: 166-167). This sounds very odd. Indeed, it is not clear why people should reason in a certain way in order to count as rational, regardless of
whether there are more efficient ways to achieve their goals. It seems to me that deontologists cannot appeal to the consequences of following the principles belonging to \( N \) in order to justify why people must follow them. If so, they lack an argument that can explain why people ought to reason systematically in accordance with a given set of normative principles rather than another one in order to count as rational.

These considerations cast some doubt on the idea that human rationality is best characterized by a deontological picture. As seen above, Cherniak has tried to accommodate this issue, that is, he has tried to tackle the question of the feasibility and effectiveness of people’s reasoning strategy within a deontological account of rationality. According to him, a rational subject ought to conform only to some norms of rationality, i.e., those which are apparently relevant to the task of inferring the more appropriate conclusions, or making the more appropriate decisions. By adopting this form of deontologism, however, one is faced with several problems in the attempt to evaluate whether or not people’s reasoning performances are normatively appropriate. When such an approach to rationality assessment is employed, one needs to specify the kind of context and goal relative to which the assessment should be made in order to select the relevant normative principle belonging to \( N \). In other words, deontologists should take into consideration several conceptual and empirical issues in order to select the appropriate normative principle when assessing a given reasoning performance and to apply it correctly. In particular, they have to identify and examine various factors before the relevant principle of rationality can be applied. However, no deontologist explains how this can be done or if there is any principled way of doing that. As a result, deontologists seem to be in trouble in characterizing what constitutes good reasoning by invoking one’s conformity to a given set of normative principles.
4. Concluding remarks

Let me summarize the main topics related to the deontological conception of rationality. In the first chapter, I have examined the normative standards which belong to the standard deontological approach to rationality assessment, i.e. the standard picture of rationality. This picture claims that there are \textit{a priori} normative principles of rationality, that these principles are universally applicable, and that they derive from deductive logic, standard probability theory, and expected utility theory. As is said by Hardy-Vallée and Thagard (2008: 176), this picture of rationality presupposes three main assumptions: (i) its normative principles are justified \textit{a priori}; (ii) they cannot be informed by empirical evidence; (iii) its normative principles are not revisable. Deontologism has been too often identified with this picture of rationality, but such an identification is misleading. As seen in this chapter, even if one assumes a deontological conception of rationality, the classical normative principles seem to be to a certain extent irrelevant in order to assess human reasoning. In particular, there are two different sorts of irrelevance: (i) the standard picture is irrelevant to human reasoning because its normative principles do not correspond to those implemented in the human cognitive system; (ii) the standard picture is irrelevant to human reasoning because its normative principles do not prescribe feasible ways to reason, nor can guide human reasoning. As we have seen, these claims are part of two different non-standard deontological accounts of rationality respectively. With regard to (i), according to Cohen, human reasoning competence does not really fit with the classical normative principles of rationality. If human reasoning falls short of the classical normative principles of rationality, there is something wrong in those principles, not in our reasoning competence. I have examined two different interpretations of Cohen’s intuition-based account of rationality: while, according to the first one, rationality amounts to the ability to conform to people’s prevailing intuitions about reasoning validity, the second characterizes
rationality as relative and calls for a comparative evaluation among different normative models based on competing ordinary intuitions about reasoning validity. Although the second interpretation provides a more satisfactory answer to the evaluation question than the first does, it calls for an analysis of the features of a situation which support the use of a particular normative model of rationality. With regard to (ii), Cherniak argues that every account of rationality should be concerned with resource-relativity. Indeed, according to him, the question of resource allocation, such as benefits and costs in using a particular reasoning strategy, plays a fundamental role in assessing human reasoning. However, I have argued that both these deontological accounts are inadequate for assessing human reasoning. As an alternative, I have considered the attempt to develop prescriptive models of rationality, that is, models of reasoning computable by human cognitive system. However, by taking this path, it is not so clear how many prescriptive models should be developed to account for people’s differences in their cognitive capacities and reasoning abilities. Finally, following Samuels and his colleagues, I have suggested that any deontological account, either standard or non-standard, is flawed as the idea that what characterizes good reasoning is conforming to a set of normative principles does not provide an explanation of why people should reason well. On this view, any deontological account of rationality has to be rejected as a possible candidate for an assessment criterion for reasoning performance and rationality.
Chapter 3 – Consequentialism and adaptation

Introduction

As seen in the previous chapter, deontologism fails in its attempt to provide a normative framework for assessing human reasoning. In the last twenty years or so, as an alternative to it, many researchers involved in the rationality debate have assumed a consequentialist conception of rationality. According to its supporters, the fact that people do not systematically reason in accordance with some alleged normative principles should not be considered as an evidence of their irrationality. It is not the compliance with an arbitrary set of normative principles that makes people’s reasoning normatively appropriate, but its effectiveness in attaining the goals people aim at. As a consequence, according to consequentialist accounts of rationality, the determination and selection of normative standards for reasoning is constrained by a specific range of goals, and reasoning performances should be evaluated by attending to their consequences. As we will see, however, different consequentialist theories focus on different sorts of consequences.

The chapter is divided into two parts. In the first part, I begin by characterizing the main assumptions and purposes of the consequentialist conception and then proceed to examine whether the standard picture of rationality can be restated and supported in consequentialist terms, concluding that it cannot.

In the second part, I examine how evolutionary psychology can be used to support the consequentialist conception of rationality. It has been generally assumed that the consequentialist approach to rationality receives a strong support from it. According to the evolutionary view of rationality, evolution, through natural selection, was directed towards
increasing reproductive success and every part of the human body, including the brain, has been selected because it is a good instrument for achieving that goal. Consequently, human reasoning is seen as having evolved to the promotion of reproductive success. However, contrary to what is held by evolutionary and ecological psychologists, I hold that this brand of consequentialism cannot give us a satisfying framework for assessing human reasoning. In particular, consequentialist accounts based on evolutionary considerations provide a very limited normative framework for assessing human reasoning as the only kinds of subject’s goal they take into account are always linked, directly or indirectly, to inclusive fitness, that is, approximately the total amount of genes that an organism transmits to future generations.

1. Consequentialism: from the standard picture of rationality to bounded rationality

I begin this section by illustrating the main assumptions and purposes of the consequentialist conception of rationality. I then examine whether the standard picture of rationality can be restated and supported in consequentialist terms. By relying on Herbert Simon’s notion of bounded rationality, I argue that any attempt to restate the standard picture in consequentialist terms is doomed to fail. Indeed, the standard picture does not seem to be able to prescribe efficient and feasible ways of attaining the relevant goals which people usually aim at. Finally, starting from Simon’s suggestions, I point out two fundamental conditions that any consequentialist approach to rationality assessment should satisfy.
1.1 Reasoning and goals

Good reasoning is such only in relation to a goal. This is the basic claim of consequentialism. In other words, according to it, reasoning is, before all, a goal-directed activity of thought. Claiming that reasoning is a goal-directed activity can be understood in two different ways: it can refer to either reasoning that is directed by goals or reasoning that is directed at goals (Nickerson 2007: 176). While, in the first sense, reasoning is taken to be a tool that people can employ to achieve a wide range of goals, which might have been chosen by the reasoner herself or set by someone else, in the second sense goals are the object of one’s reasoning (for example, one can reason about whether her goals are consistent with her preferences or value, or, when faced with conflicting goals, which she should try to attain). Consequentialists adopt the first way of conceiving of reasoning as a goal-directed activity, which clearly presupposes an instrumentalist view of rationality, according to which human rationality is concerned with how to reach one’s goals rather than with what goals are worth achieving.

Consequentialism is generally characterized in opposition to deontologism. Indeed, according to consequentialists, what characterizes good reasoning is not compliance with a given set of normative principles, but effectiveness in attaining some specific goals such as personal utility, reproductive success, or happiness. Since goals can be characterized in naturalistic or personal terms, as having an extrinsic or intrinsic value, and so on, consequentialist accounts of rationality differentiate each other as to the kind of goals at which good reasoning should aim (see Hardy-Vallée & Thagard 2008: 177). Conforming to normative principles can be useful when it contributes to attain the goals people aim at, but it is not constitutive of good reasoning: “to reason correctly is to reason in such a way that a particular reasoner is likely to attain certain goals or outcomes” (Samuels et al. 2004: 166). Given a specific range of relevant goals or outcomes, consequentialist approaches to
rationality assessment evaluate reasoning performances in terms of their effectiveness and efficiency in achieving some specific range of goals or outcomes.

Which implications does such a conception have for what has been called the normative and the evaluative question? As seen in the first chapter (see Chapter 1, Section 1.1), normative models of rationality are supposed to provide standards of rationality. From a consequentialist point of view, a normative model is relevant to a given reasoner only insofar as following it helps her reach her goals. Accordingly, the characterization of the rationality condition for individuals requires deep changes. The general condition for rationality defines what is required for an individual to be rational. In deontological accounts of rationality, their corresponding rationality conditions articulate the connection between the rationality of an individual and her capacity to reason, implicitly or explicitly, in accordance with the relevant norms of rationality (see Chapter 1, Section 2.4; Chapter 2, Section 3.1). In contrast, in consequentialist terms, the rationality condition can be characterized as follows:

(R3) Given a specific range of goals $G$, a subject $S$ qualifies as rational if and only if, in order to attain one of the elements of $G$, $S$ relies on the most efficient and effective means.

Much as (R1) requires a characterization of the set of norms of rationality, requirement (R3) presupposes the characterization of the range of goals which good reasoning should serve. The fact that consequentialist accounts focus on the importance of the outcomes of reasoning has an important implication. As is observed by Samuels and his colleagues, in contrast with deontological accounts, consequentialist accounts satisfy a basic condition for any normative theory of reasoning, that is, the value condition: “a normative theory of reasoning should provide us with a vindication of rationality. It should explain why reasoning in a normatively correct fashion matters – why good reasoning is desirable”
On a consequentialist view, good reasoning matters because it is a good instrument to attain one’s goals. But, if the goal of a given subject $S$ is to maximise her true beliefs and expected utilities, could we restate the standard picture of rationality in consequentialist terms? Could its normative principles be interpreted in consequentialist terms? As Jonathan Baron has pointed out (2008: 61), according to a consequentialist view, an appropriate (and therefore rational) thinking is

[...] whatever kind of thinking best helps people achieve their goals. If it should turn out that following the rules of formal logic leads to eternal happiness, then it is “rational thinking” to follow the laws of logic (assuming that we all want eternal happiness). If it should turn out, on the other hand, that carefully violating the laws of logic at every turn leads to eternal happiness, then it is these violations that we shall call “rational”.

1.2 Consequentialism and the standard picture of rationality

According to the standard picture of rationality, to be rational means to reason (explicitly or implicitly) in accordance with a specific set of normative principles. As is noted by Hardy-Vallée and Thagard (2008: 177), in the standard picture “effectiveness is taken for granted but is not evaluated”, that is, it is supposed that a reasoner that conforms to its normative principles will have a better life than another one who systematically deviates from them, “yet this assumption is never subject to any empirical verification”. As seen in the first chapter (see Chapter 1, Section 2.2), the standard picture is supposed to have two main virtues, namely the necessity and universality of its principles. Consider, for example, the normative principles of rationality derived from expected utility theory: (i) if one follows them, she will always maximize her expected utility and (ii) they should work equally well in any context. If so, following these principles is the most effective way of
attaining one’s goals in every context. That is in principle true. In my view, however, maximization (and optimization) requires what Carol Rovane (2004: 322) calls “all-things-considered” judgments, that is, judgments about what is the best thing to do given all one’s beliefs and preferences. In a consequentialist characterization of the standard picture of rationality, what people ought to do in order to count as rational is to ground their thinking and actions on all-things-considered judgments. As is pointed out by Rovane (2004: 323), these judgments are “the outcome of a variety of rational activities that together comprise one’s decision”, such as removing contradictions among one’s beliefs and making order of one’s preferences transitively, all of which are activities directed at meeting specific normative standards of rationality, and are concerned with consistency and transitivity of preferences respectively. By assuming a consequentialist perspective, one holds that all these rational activities can be characterized as having a common purpose, so that they together should lead to the attainment of one’s goals. Consider now what would happen if people decide what to do without conforming to the relevant normative principles of rationality. As Rovane suggests (2004: 323), if a subject decided to not meet the requirement of consistency, by persevering in believing a contradiction, then it will impossible for him to figure out what it is the best thing to do given all her beliefs and preferences. While one belief might prompt the subject to performing a certain action, another one might direct her to refraining from it. Similarly, not following other basic norms of rationality, such as the requirement of the transitivity of preferences, will preclude the possibility to do the best thing that one can do and so maximise her expected utility. As a result, in order to attain one’s goals, it seems to be a necessary requirement to ground our actions on the basis of all-things-considered judgments.

Although the argument supporting the consequentialist account of the standard picture of rationality seems to be compelling, it relies on a questionable assumption, that is, it
assumes that when examining people’s decisions and inferences we are allowed to ignore the resource-limited nature of the human mind. If people had unlimited cognitive resources, they would be able to make all-things-considered judgements and would always entertain consistent beliefs and coherent and ordered preferences. But such ideal people do not exist in reality: people’s cognitive limitations should be taken into consideration when considering the extent to which they can maximize (or optimize) the goals they aim at. Therefore, any attempt to restate the standard picture in consequentialist terms is doomed to fail.

1.3 Bounded rationality

About fifty years ago, Herbert Simon coined the label “bounded rationality” in his Models of Man (Simon 1957: 198). Bounded rationality is considered to be a benchmark for all who want to develop a consequentialist account of rationality taking into account people’s cognitive limitations. In the last three decades, the expression “bounded rationality” has become a term used to designate different and sometimes competing views of reasoning and decision-making. It is often used in different ways even within the same discipline, i.e., in order to criticise the classical normative models of rationality, to develop behavioral models, to articulate concepts of rationality and so on (see Gigerenzer & Selten 2001b; Grüne-Yanoff 2007). However, the general idea underlying this notion is that, when people reason or make decisions, they consider only a limited number of alternatives, that is, the options that are good enough given their cognitive and environmental constraints.

1.3.1 Maximizing or satisficing?

Working on decision making, Simon showed that, while expected utility theory requires very complex calculations and complete information in order to single out the decision that
maximises one’s expected utility, people have too limited cognitive resources and usually lack the basic information needed to solve the decision problem they are faced with. As a result, it would be impossible for them to perfectly work out the outcomes of the alternative options among which they have to choose. Being things so, Simon held that optimization cannot be our ultimate goal when making decisions or performing other type of human behaviour. Specifically, according to him, people reason and make decisions under environmental and cognitive constraints, such as limited memory and the need to reduce information costs.\(^1\) Simon proposed an analogy to explain this point: “human rational behaviour […] is shaped by a scissors whose two blades are the structure of the task environments and the computational capabilities of the actor” (Simon 1990: 7). According to him, if you want to understand how the human mind works, you have to look at both blades of the scissor (see also Gigerenzer & Selten 2001b). Because of these constraints, in order to count as rational a subject ought to make *satisficing* choices; when faced with a decision problem, she ought not to seek optimal or maximizing solutions, but only acceptable solutions. As Simon puts it, a rational decision maker ought not to try to maximize her expected utility because that is often an impossible task: in ordinary life, all of us usually behave as “satisficers”. For example, if decision makers recognize and choose the type of response appropriate to the circumstance in which the choice is being made, *satisficing* avoids unsatisfactory outcomes and reduces cognitive effort. Within this normative framework, rationality requires of decision makers that they should search for information and process it in ways that are less expensive than classical models of rationality require (Simon 1976: 136).

\(^1\) In subsequent research, the first aspect has received an overwhelming amount of attention. But recently, as we will see, the interest for the second aspect is widely increased.
1.3.2 Is bounded rationality really normative?

In the fifties, Simon wondered whether the “[…] model of economic man provides a suitable foundation on which to erect a theory – whether it be a theory of how firms do behave, or of how they ‘should’ rationally behave” (Simon 1955: 99). Likewise, Gigerenzer and Selten (2001b: 6) have pointed out with respect to Simon’s view of rationality that “bounded rationality means rethinking the norms as well as studying the actual behaviour of minds […]”. Understood in these terms, Simon’s proposal has to be considered as an alternative to the classical normative models of rationality, not an account that accepts them and analyzes whether and to what extent reasoning performances deviate from their normative principles. The real question is then whether Simon’s proposal succeeds in providing an appropriate normative framework for assessing human reasoning. As is noted by Nickerson (2007: 21), “a decision need be neither optimal nor even as close to optimal as one is capable of achieving in order to meet the dictates of rationality. How close to optimal a decision must be to be considered rational appears to be an open question”. On the one hand, any deviations from the classical normative standards of rationality may always be reconsidered as being reasonable enough, given some cognitive or environmental constraints that were previously disregarded. On the other hand, however, “there are so many ways in which rationality can be bounded that we can never be sure we have the right one” (Watts 2003: 66; as reported by Nickerson 2007: 358). In order to determine whether satisfying choices are good enough for a certain individual, an evaluator must know something about her, such as what her goals are and what her level of aspiration is: while having this information is sometimes possible, it is often too difficult a task for anyone. Moreover, as is pointed out by Edmund Henden (2007), it is wrong to think that Simon and his followers hold that it is always rational to make satisfying choices, i.e., prefer good enough choices to optimal ones. This seems to be very counter-intuitive. A more plausible
way to understand their claim is to assume that, according to the supporters of bounded rationality, it is not always rational to aim at *satisficing* choices rather than optimal ones. As Henden argues (2007: 341), “this weaker claim is consistent with the possibility that there may be circumstances in which genuine satisficing in fact would be irrational”. Consequently, he goes on, pointing out that showing few examples in which being a *satisficer* is not rational is not sufficient to reject *satisficing*. Again, however, we should conclude that proponents of bounded rationality need an argument about what conditions need to be present in order for a case of *satisficing* to be rational (Henden 2007: 341-342).

### 1.4 From maximization to satisficing: a final balance

The comparison between the *standard picture* and the concept of bounded rationality provides useful suggestions for the development of any satisfactory consequentialist account of rationality. In particular, I identify two fundamental conditions that any consequentialist approach to rationality assessment should satisfy.

First, any consequentialist account of rationality that should be used as a benchmark against which to evaluate human reasoning must take human cognitive limitations into account. On the one hand, this condition focuses on the issue of the limitations of human rationality. On the other hand, however, it also assumes that by taking these limitations into account we can reach an adequate notion of rationality. As seen above, behaviour that is optimal in terms of the classical normative standards is seldom required for achieving people’s realistic, ordinary goals. According to Simon, maximization or optimization is often an impossible task because the calculations required in maximizing one’s utility or in achieving optimal solutions are usually computationally intractable. If so, the *standard picture of rationality* seems to require of humans to make inferences that are far beyond their computational capacities.
There is then a second condition which has important implications for how we should think about normative standards of rationality according to a consequentialist view: consequentialism should reject “the assumption that the same standards of good reasoning apply in all environments – that they are context invariant” (Samuels et al. 2004: 170). Clearly, different contexts can affect the efficiency of a reasoning strategy in different ways. Accordingly, the second condition states that an appropriate consequentialist account of rationality should index standards of evaluation to the environment in which reasoning takes place.

According to Simon (1981; 1990), these two requirements could be satisfied by developing a science of adaptive systems. By the expression “adaptive system”, Simon means “a system resulting from his reactions to modelling forces of the environment to which every system must be adapted in order to survive” (Simon 1990: 3). Following similar considerations, starting with the late eighties the notions of adaptation and environment have received increasing attention in the works of philosophers of mind, cognitive psychologists and anthropologists.

2. Rationality and adaptation

Debating of human rationality, a question arises: if our capacity to reason evolved through natural selection, why was it selected? That is not likely to have happened thanks to some intrinsic value of that capacity. As is said by David Over (2004: 9), from an evolutionary standpoint the human capacity to reason is as much valuable as the long neck of a giraffe: both characteristics were selected because they were good instruments for increasing the reproductive success of humans and giraffe respectively. In particular, with regard to our evolutionary ancestors, increasing ability to reason helped them “to communicate, operate,
and plan more effectively, allowing them to get more food and facilitating successful reproduction in other ways” (Over 2004: 9). On this view, human reasoning is seen as having evolved to the promotion, direct or indirect, of reproductive success.

In this section, I review two different kinds of evolutionary arguments: (i) the former assumes that every cognitive adaptation is truth-conducive; (ii) the latter is concerned with the development in human mind of an indefinite number of modules which are useful for solving problems that our ancestors were faced with.

2.1 Adaptations as truth-conducive

According to argument (i), evolution necessarily favours organisms that form true beliefs and thereby reason rationally (e.g. Dennett 1981; 1987; Fodor 1981). The evolutionary argument for rationality usually proceeds in two steps. First, it is argued that evolution, through natural selection, will favour cognitive mechanisms that generate a preponderance of true beliefs over false ones. Second, it is assumed that those cognitive mechanisms that reliably generate true beliefs should be considered as rational cognitive mechanisms. Therefore, it is concluded that evolution will favour organisms which possess rational cognitive mechanisms and thereby are themselves rational. As is noted by Edward Stein (1996: 173), “the two-step evolutionary argument moves from a connection between evolution and truth on the one hand and a connection between truth and rationality on the other, to a connection between evolution and rationality”. The first connection is based on the idea that having true beliefs is more adaptive than having false ones. Accordingly, an organism that reliably generates true beliefs has a better chance to survive and reproduce than an organism that tends to generate false beliefs. As Dennett (1981: 17) has observed, “the capacity to believe would have no survival value unless it were a capacity to believe

---

2 A more detailed reconstruction of the evolutionary argument is to be found in Stich (1990: 56-59) and Stein (1996: 174-213) (for an exposition and discussion of the argument see also Pacherie 2002: 309-311).
truths”. On this view, if natural selection favours organisms equipped with cognitive mechanisms that are good at generating true beliefs and avoiding falsehoods, the fact that humans have not yet become extinct demonstrates that their belief-generating cognitive mechanisms do a good job at generating truths. Assuming that an organism possessing cognitive mechanisms which reliably generate true beliefs is rational, Daniel Dennett (1987: 33) concludes that “evolution has designed human beings to be rational, to believe what they ought to believe and what they ought to want”. On this view, the evolutionary argument provides a good explanation as to why humans should be considered as rational.

As many biologists and also philosophers have argued, however, the evolutionary argument attributes a function to natural selection that seems to be very questionable (see, e.g., Lewontin 1995; Stein 1996; Stich 1990). We have no basis for claiming that natural selection should select organisms which are by and large rational in the sense that their belief-generating cognitive mechanisms reliably lead to true beliefs. Stephen Stich has offered two famous arguments in favour of this objection (Stich 1990: 55-74). According to these two arguments, evolution, through natural selection, would not necessarily favour organisms equipped with cognitive mechanisms that are good at producing true beliefs. The first argument has to do with internal fitness, which Stich characterizes as a function of the ratio between the epistemic benefits obtained from possessing an information and the resources required to obtain it. According to this argument, evolution, through natural selection, might have selected a less reliable belief-generating cognitive mechanism over a more reliable one as the former might have led to a better level of internal fitness. In other words, the more reliable belief-generating cognitive mechanism might have costs that are higher than its potential epistemic benefits. The second argument is concerned with external fitness, that is, fitness in relation to the environment in which our ancestors evolved (Stich 1990: 60; see also Sober 1981: 105). Stich (1990: 62) argues that
a very cautious risk-aversive inferential strategy – one that leaps to the conclusion that danger is present on very slight evidence – will typically lead to false beliefs more often, and true ones less often, than a less hair-trigger one that waits for more evidence before rendering a judgment. Nonetheless, the unreliable, error-prone, risk-aversive strategy may well be favoured by natural selection. For natural selection does not care about truth; it cares only about reproductive success. And from the point of view of reproductive success, it is often better to be safe (and wrong) than sorry.

So, according to Stich, the evolutionary argument fails as natural selection cannot guarantee that the cognitive mechanisms it selects are good at producing true beliefs. As a consequence, there is no reason why evolution, through natural selection, should have selected organisms which are by and large rational. However, we can substitute truth with reproductive success and so put forward a different kind of connection between rationality and evolution. Indeed, although natural selection does not select cognitive mechanisms that are good at producing true beliefs, it reliably selects cognitive mechanisms that increase our reproductive success. This idea is endorsed by evolutionary psychologists.

### 2.2 Evolutionary psychology

Recently, David Buller (2005) has argued that there are two different ways of understanding evolutionary psychology. According to the former, “it is simply the study of human behaviour and psychology from an evolutionary perspective” (Buller 2005: 277). That interpretation is not relevant for our current issue. The latter (to be found in works by Leda Cosmides, John Tooby, Steve Pinker and David Buss) understands evolutionary psychology as an evolutionary approach to human behaviour which “entails specific doctrines regarding the nature and evolution of the human mind” (Buller 2005: 277). While
the classical evolutionary approach to the human mind aims at discovering the evolutionary origins of cognitive mechanisms, this competing approach has as its main goal that of discovering human psychological adaptations which, it is argued, are the “proximate mechanisms” that cause human behaviour (e.g. Cosmides & Tooby 1987; 1994). According to evolutionary psychologists, the current human mind consists of these psychological adaptations, which are characterized as domain-specific cognitive mechanisms that are known as “Darwinian modules”. These modules were selected by natural selection because they do a good job at solving adaptive problems, that is, any “evolutionary recurrent problem whose solution promoted reproduction, however long or indirect the chain by which it did so” (Cosmides & Tooby 1994: 87). Every module is dedicated to solving a restricted set of problems in a restricted domain. If one module works well for solving a given adaptive problem belonging to one domain, it will not work equally well if applied to another similar adaptive problem but belonging to a different domain. So, every domain-specific cognitive mechanism is claimed to be caused by a different adaptive problem which our ancestors were frequently faced with.

But how do evolutionary psychologists characterize those domains? As is said by Cosmides (1989: 194n), “our cognitive mechanisms should be adapted to the hunter-gatherer mode of life, and not to the twentieth century industrialized world”: they should have been selected to solve adaptive problems that our hunter-gatherer ancestors met in the Pleistocene, which corresponds to the period included between 1.8 million to 10.000 years ago (see Tooby & Cosmides 1992). If the current human mind consists of cognitive modules that are geared to solve adaptive problems our Pleistocene ancestors were frequently faced with and which had a high impact on their survival and reproductive success, those modules should work well whenever they receive information in a format similar to that which was available in the so-called “environment of evolutionary
adaptation” (EEA). This environment is defined then as “a statistical composite of all the different environments in which our human, hominid, primate, and mammalian ancestors lived” (Brase 2004: 317).

2.3 Reasoning, rationality and adaptation

In the last two decades, evolutionary psychologists, such as Leda Cosmides and John Tooby, have provided both theoretical considerations and empirical data in order to show that the most common subjects’ responses in the classical reasoning experiments should be considered as normatively appropriate after having taken into account the evolutionary conditions in which the cognitive mechanisms providing those responses have evolved (e.g., Cosmides 1989; Cosmides & Tooby 1996; Fiddick et al. 2000). Their analysis is intended to show that supporters of the Irrationality thesis, such as researches in the heuristics and biases tradition, have not demonstrated that people are deeply irrational. As Samuels and his colleagues (2002: 244) put it, according to evolutionary psychologists, “human reasoning is not subserved by ‘fast and dirty’ heuristics but by ‘elegant machines’ that were designed and refined by natural selection over millions of years”. As a result, in their view, people’s reasoning is the result of the processing performed by the cognitive modules that were evolved to solve the evolutionary problems our ancestors were frequently faced with, and so they draw normatively appropriate conclusions only when the information of a reasoning problem is given in the appropriate format and the reasoning problem itself resembles a problem from the environment of evolutionary adaptation. On this view, good reasoning and decision making should lead to successful outcomes but, at the same time, from an evolutionary perspective, most evolutionary psychologists define success as linked to inclusive fitness, that is, in terms of the number of genes one transmits to subsequent generations as a result of direct and indirect reproduction (Hamilton 1963),
arguing that behaviour is rational when it increases inclusive fitness. In other words, the evolutionary thesis about human rationality supported by evolutionary psychologists is based on the assumption that humans, we as well as our ancestors, tend to behave in ways that increase their inclusive fitness. Since evaluations of reasoning performances are made by looking at the contributions of their outcomes to inclusive fitness, human rationality is not assessed against short but rather long-term criteria. The evidence that humans are rational amounts to the fact that they have evolved in such a way as to be able to follow “Darwinian” algorithms which enable them to handle recurrent adaptive problems confronted during the Pleistocene. As is pointed out by Samuels and his colleagues (2002: 256), evolutionary psychologists implicitly assume the distinction between the actual domain and the proper domain for a mental module, as proposed by Dan Sperber (1994). Sperber has illustrated how a given mental module can be expected to work in terms of its scope of application. First of all, there is the proper domain of a mental module. The proper domain amounts to the class of situations that recall the specific adaptive problem which a mental module evolved to solve. That is, the proper domain for a mental module is “all the information that it is the module’s biological function to process” (Sperber 1994: 52). However, many domain-specific cognitive mechanisms can be triggered in situations that are not actually within their proper domains. This is what Sperber calls the actual domain for a mental module. In Sperber’s words, the actual domain for a given mental module amounts to “all the information in the organism’s environment that may (once processed by perceptual modules, and possibly by other conceptual modules) satisfy the module’s input conditions” (Sperber 1994: 51-52). Evolutionary psychologists assume that, when evaluating reasoning performances, assessments should be always relativized to the proper domain of the cognitive mechanisms that have produced the performances. This means that
human rationality is defined by reference to the specialized cognitive mechanisms that have evolved for solving specific domains of reasoning (see also Samuels et al. 2002: 256).

Consider the implications that such an evolutionary view has for the interpretation of the responses in two well-known reasoning tasks which I have introduced in the first chapter (see Chapter 1, Section 3.1). First, Cosmides and Tooby claim to possess evidence for the existence of a domain-specific cognitive mechanism that activates in experiments based on the selection task (e.g. Cosmides 1989; Cosmides & Tooby 1992). That task has been studied using different kinds of conditional statements, some of which elicit the correct response in most subjects’ reasoning performances. In particular, subjects improve their performances when the task they are given has more realistic content, such as in the following case:

If a person is drinking beer \(p\), then that person must be over 19 years of age \(\neg q\).

(Griggs & Cox 1982: 415)

In this modified version of the selection task, subjects are required to turn over the cards which might show whether the conditional rule cited above has been violated. When such a task is given to subjects, most of them make the right choice, that is, they select the \(p\) card, which has printed on its visible side “drinking beer”, and the \(\neg q\) card, which has printed on its visible side “15 years old” (Griggs & Cox 1982: 414-417). In order to explain why only certain modified versions of the selection task elicit the correct response, Cosmides and Tooby have proposed the Cheater-Detection Hypothesis (Cosmides 1989; Cosmides & Tooby 1992). According to their hypothesis, humans possess a module devoted to solving problems about social exchange, which they characterize as “adaptive cooperation between two or more individuals for mutual benefit […]” (Cosmides 1989: 193). In particular, a

3 More generally, the drinking age problem has been characterized as a deontic version of the selection task, as opposed to the standard indicative version of it (for a discussion on deontic versions of the selection task see, e.g., Manktelow & Over 1991).
recurrent adaptive problem our evolutionary ancestors would face with was to detect whether their conspecifics were trying to cheat them when involved in social exchanges. In Cosmides and Tooby’s view, we should have inherited from our evolutionary ancestors a module devoted to detecting people cheating on social contracts. This domain-specific module evaluates whether a social contract is upheld or violated, and in the selection task it prompts subjects to turn the cards that are most informative for determining this. Roughly speaking, a social exchange is characterized as an exchange between two individuals that is regulated by the following conditional rule “If you take the benefit $A$, then you must pay cost $B$”. A violation of such an arrangement, that is cheating, takes place when someone gets a benefit without paying the cost, which is a violation of a social agreement as opposed to a violation of logic. According to the Cheater-Detection Hypothesis, this domain specific module can be applied to the task concerning the drinking-age rule but not to the standard abstract version, as the drinking-age rule can be seen as a social contract and so violating it can be regarded as cheating. If the conditional statement involves a social contract subjects will react in a predetermined way, such as in the case of the drinking-age rule; the implicit reference to a social contract affects what goal they form. In particular, selection framed in social-contract terms where cheating can be seen as a risk will activate the cheater-detection module within its proper domain, which produces the correct response. Selection tasks that do not have this form, such as the standard abstract version, would not activate the module.4

Another analysis is concerned with the format in which statistical information is presented. Turning to probability judgments, Cosmides and Tooby have proposed the so-called Frequentist Hypothesis, according to which “some of our inductive reasoning mechanisms do embody aspects of a calculus of probability, but they are designed to take frequency information as input and produce frequencies as output” (Cosmides & Tooby

---

4 This hypothesis, however, remains highly controversial. For alternative explanations see, e.g., Cheng & Holyoak 1985; Oaksford & Chater 1994; Sperber et al. 1995.
In other words, they hold that in the environment of evolutionary adaptation, statistical information was available to our evolutionary ancestors as frequencies rather than as probabilities, such as “four out of every ten times you go into this forest you meet a lion” rather than “the probability of meeting a lion going into this forest is 0.4” (Cosmides & Tooby 1996). This should explain why human subjects usually are not so good at reasoning with probabilities of single events, but make good performances when reasoning about frequencies. As a matter of fact, most empirical studies on probability judgments have demonstrated that a great majority of participants respond correctly when statistical information is presented as frequency (see Fiedler 1988; Hertwig & Gigerenzer 1999). In the “frequentist” version of the Linda Problem devised by Klaus Fiedler (1988) (for the standard version see Chapter 1, Section 3.1.2), for example, the description of Linda is followed by a question like this:

To how many out of 100 people who are like Linda do the following statements apply?

Linda is a bank teller.

Linda is a bank teller and active in the feminist movement.

While in the standard version of the task 85% of the subjects gave the wrong answer, only few subjects (about 13%) did the same in the “frequentist” version (Fiedler 1988; see also Cosmides & Tooby 1996: 17-21).

3. Ecological Rationality

In working on cognitive adaptations and their relation with rationality, a further step should be to identify the relevant features of the domain for which a particular cognitive
mechanism was selected and then proceed to decide in which other actual domains, displaying similar characteristics, the cognitive mechanism may still operate. Such an analysis has been developed by Gerd Gigerenzer and the ABC group. They have amended to a certain extent the hypothesis proposed by evolutionary psychologists such as Leda Cosmides and John Tooby and have gone one step further (see mainly Gigerenzer 2000; Gigerenzer 2008; Gigerenzer & Selten 2001a; Gigerenzer et al. 1999). They propose a different view on the rationality question, according to which one needs ecological standards for assessing human reasoning. While evolutionary psychologists hold that evolution has endowed humans with an indefinite number of domain-specific cognitive mechanisms, Gigerenzer and his collaborators maintain that humans are equipped with “an adaptive toolbox” of fast and frugal heuristics. These heuristics make them, not only adaptively, but also ecologically rational.

3.1 Ecological validity

Gerd Gigerenzer and his colleagues introduce their proposal by appealing to three different forms of rationality, that is, bounded, ecological, and social rationality, and arguing that only an analysis of the relationship among these three aspects of rationality can lead to a comprehensive understanding of what it amounts to. By relying on these three forms of rationality, they hold that we can understand human behaviour and cognition in their adaptation to specific environments and discover the heuristics that guide them (Gigerenzer & Todd 1999: 25). In particular, a leading role is attributed to the notion of ecological rationality, which is characterized as rationality adapted to the environment. Indeed, behind this proposal, there is the idea that “the success of boundedly rational heuristics depends on their ability to exploit the information structures in the ecological and social environment” (Todd & Gigerenzer 1999: 360). More specifically, Gigerenzer holds that a heuristic is
“ecologically valid” when it is successfully applied in an environment that has a relatively fixed and homogeneous structure: the rationality of heuristics does not amount to logicality, but rather is ecological (Gigerenzer & Todd 1999: 18-19). In other words, a heuristic process, as well as any adaptation, can be evaluated as rational or irrational only in relation to an environment. In particular, Gigerenzer and Goldstein (1996: 665-666) have argued that ecological standards of rationality should replace classical ones as the appropriate benchmark against which to assess reasoning performances, and these standards should be based on: (a) the relevance of the environment; (b) the recognition of cognitive limitations; and (c) the identification of specific algorithms existing in the human mind.

3.2 The adaptive toolbox

At a cognitive level, the ecological approach to rationality assumes that humans have at their disposal an adaptive toolbox (see, e.g., Gigerenzer 2001; Gigerenzer & Todd 1999). It consists of a set of domain-specific cognitive mechanisms, which can be inherited, learned or designed: just as screwdriver and hammers are designed for specific classes of activity, so are the cognitive mechanisms belonging to the adaptive toolbox. In characterizing those cognitive mechanisms, Gerd Gigerenzer (2001: 6) specifies that they include both Simon’s satisficing and what he calls fast and frugal heuristics. While the former involves “search across alternatives […] assuming that a criterion is given”, fast and frugal heuristics work in the opposite way, that is, they “search for criteria or cues, assuming that the alternatives are given” (Gigerenzer 2001: 121). According to Gigerenzer and his collaborators, heuristics are “fast” because they require simple computations and “frugal” because they need little information to come to a conclusion; so, they are considered to be very effective in environments in which people have limited time and

---

5 It is “adaptive” because it is a product of natural selection, and is a “toolbox” because it is composed by a number of domain-specific cognitive tools that can be used in a flexible way (Gigerenzer & Todd 1999).
information for making decisions and drawing inferences (Gigerenzer & Todd 1999: 24). Their frugality, however, does not compromise their effectiveness. Indeed, if fast and frugal heuristics are applied to the right kind of environment, they are very effective and lead reliably to good solutions. Gigerenzer and his collaborators hold that, in order to explain how a given heuristics works, one needs to identify (1) its building blocks (such as rules for search, stopping search, and drawing conclusions), and (2) the relevant characteristics of the environments in which that heuristic works well (e.g., Gigerenzer 2001: 501-502; Gigerenzer & Todd 1999: 24-25).

By way of illustration, consider, for example, the “recognition heuristic”. Suppose that you are asked to decide which of two objects has a higher value on some criterion. Gigerenzer and Goldstein (1999) have empirically demonstrated that, in order to make such a decision, people are naturally inclined to follow the following rule: “if one of two objects is recognized and the other is not, then infer that the recognized object has the higher value” (Gigerenzer & Goldstein 1999: 41). This is what they call the “recognition heuristic”. In particular, how the recognition heuristic works could be illustrated by referring to an experimental case in which lack of information increases the success of a prediction (Goldstein & Gigerenzer 1999: 43). In this experiment, two groups of subjects, respectively composed by German and American college students, were given the name of two American cities and then asked which has a larger population:

Which U.S. city has more inhabitants: San Diego or San Antonio?

While 100% of German students answered correctly (San Diego), only 62% of American students did so. According to Goldstein and Gigerenzer, this surprisingly large difference is due to the fact that German students, who do not know so much about medium-sized American cities than American students, are unlikely to have ever heard of San Antonio, and thereby chose the city which they immediately recognize. This immediate choice
without any calculation seems to be appropriate in this and other similar cases because it is more likely that a foreign city one recognizes has more inhabitants than one that he does not. So, reliance on the heuristic that guides subjects to choose the city they immediately recognize has been shown to be useful. However, it would be an error to think that the recognition heuristic is completely reliable because its success depends on whether or not some particular contingent conditions occur. In particular, when relying on the recognition heuristic, one must know neither too much nor too little about the target domain. In the case mentioned above, the American students cannot apply the recognition heuristic because they know too much about American cities. Because of their familiarity with American cities, the recognition heuristic cannot be properly used as a filter in deciding which of them has more inhabitants. Similarly, if the German students had known nothing about American geography, they would have not been in a position to apply the heuristic.

### 3.3 Ecological rationality and normativity

Gigerenzer and his collaborators hold that classical principles of rationality, which are justified *a priori*, when brought to bear on real-world activities yield inappropriate standards of rationality and prove unecological. In their view, reasoning should be always evaluated in terms of its success in dealing with relevant problems which are situated in a natural or social environment (e.g., Gigerenzer & Todd 1999: 21-22). For example, although the recognition heuristic cannot be rationally justified, it usually leads to good outcomes when applied in an appropriate environment. So, in order to judge people’s reasoning performances, we should rely on an ecological standard of rationality according to which good reasoning will be always environmentally situated. Gigerenzer and his collaborators hold that these standards correspond to “multiple correspondence criteria relating to real-world”: reasoning performances must be compared with the actual
requirements of their environment, which include “making accurate decisions, in a minimal amount of time, and using a minimal amount of information” (Gigerenzer & Todd 1999: 22). In their view, those criteria should receive different weights in relation to the environment in which reasoning performances occur; in some cases, for example, making a quick decision can be more important than focusing on its accuracy and this must influence how the final decision is evaluated. Indeed, “[t]he function of heuristics is not to be coherent. Rather, their function is to make reasonable, adaptive inferences about the real social and physical world given limited time and knowledge” (Gigerenzer & Todd 1999: 22). So, every assessment should be constrained by the environment and people should not be asked to solve problems which are contextualized in environments that are not so relevant for them. On the contrary, according to Gigerenzer (1991; 1996), in the reasoning tasks created by Kahneman and Tversky subjects are located into artificial experimental contexts and receive information to which they are not ecologically adapted.⁶ Both the researchers of the “heuristics and biases” approach and Gigerenzer assume that people are naturally inclined to rely on simple heuristics but, from that, they draw different conclusions: according to Kahneman and Tversky, these heuristics are less reliable than classical principles derived from deductive logic, standard probability theory and expected utility theory, while, according to Gigerenzer, when applied in the appropriate context heuristics are about as reliable as the classical normative principles (for an interesting discussion, see Samuels et al. 2002).

⁶ See also Gigerenzer 1991; Gigerenzer & Hoffrage 1995; Gigerenzer & Hug 1992; Hertwig & Gigerenzer 1999.
4. The implications of adaptive and ecological rationality for the evaluative question

As seen in the first section (see Chapter 3, Section 1.3), Simon argued that human rationality is bounded by cognitive and environmental constraints. While traditionally these constraints have been regarded as independent, leading to very partial notions of bounded rationality, adaptive and ecological rationality put these constraints together to form a single set of constraints. The proponents of these approaches may provide a satisfactory consequentialist view of human reasoning, which takes into serious consideration the role of cognitive and situational constraints. In my view, however, there are two main objections to this kind of approach. First of all, they provide a very limited normative framework for assessing human reasoning as the only kinds of subject’s goal they take into account are always linked, directly or indirectly, to inclusive fitness. Secondly, both evolutionary psychologists and Gigerenzer justify the normative appropriateness of a reasoning performance by referring to the environment of evolutionary adaptation where its underlying cognitive mechanisms have evolved, rather than to its current context.

4.1 Adaptive rationality and context of evaluation

Evolutionary psychologists hold that whenever classical normative principles of rationality and actual reasoning strategies do not agree, we should prefer the latter to the former. According to them, there are no other standards than those which come from psychological capacities. On this view, a cognitive mechanism that operates normally is operating as it ought to operate. More specifically, the “ought” should always be understood in evolutionary terms, in the sense that a properly operating module or heuristic should be regarded as solving the problems for which it has evolved: such is the fundamental form of normativity. In other words, evolutionary psychologists maintain that natural selection has equipped humans with well designed cognitive modules that activate normatively
appropriate reasoning strategies in the kinds of problems that were relevant for them in the environment of evolutionary adaptation. So, according to evolutionary psychologists, the evolutionary history of domain-specific cognitive mechanisms makes it possible to explain why an individual appears to have a particular goal or use a certain reasoning strategy. Within this evolutionary framework, evolutionary psychologists hold that the contextual validity of reasoning strategies is always relative to their proper domains, i.e., the environments of evolutionary adaptation. However, humans modify their social and natural environment in ways that natural selection cannot follow and so nothing guarantees that reasoning strategies which were well adapted to the social and natural environments of our evolutionary ancestors provide correct answers when reasoning about contemporary problems. What happens, for example, with all of the tasks we are faced with that do not have any selective impact? As is pointed out by Stanovich and West (2003: 203), evolutionary psychologists try to set aside the differences between the environment of evolutionary adaptation and the modern environment by holding that “evolution insures a mesh between the principles of the mind and the regularities of the world” (Cosmides & Tooby 1992: 72; as reported by Stanovich & West 2003: 203). In other words, they assume that it is possible to find similarities between the environment of evolutionary adaptation and the current world.

Consider, however, an example mentioned above. Cosmides and Tooby hold that the human mind is designed to solve problems that have statistical information expressed in a frequency format. This hypothesis has been confirmed by experimental studies based on a revised version of the Linda problem in which subjects are asked to reason about frequencies rather than probabilities of single events: in such revised versions of the Linda problem subjects’ performances drastically improve. That is true. However, as is pointed out by Stanovich (2004: 135), Cosmides and Tooby’s findings do not remove “the necessity
of being able to process probabilistic information when it is presented in the current world” (Stanovich 2004: 135). In my view, their strategy of defending human rationality has too high a cost because the range of reasoning performances that evolutionary psychologists can assess becomes limited. Indeed, it is not possible to come up with an appropriate evolutionary account for all reasoning tasks in the current world. If the reasoning problems of our modern society are the same as those represented in classical reasoning experiments, and thereby they substantially differ from the evolutionary problems our ancestors were frequently faced with, evolutionary psychologists should accept that current humans behave as irrationally as they have appeared to do in the classical experiments on human reasoning. According to Stanovich (2004: 135), evolutionary psychologists can be said to be guilty of implicitly assuming that if the human mind can be shown to have adapted to certain problems (other than those posed by modern society) during evolution, then that is enough to hold that there is no cognitive problem with the subject’s failures in reasoning performance. As is remarked by Papineau (2003: 56-57),

the evolutionary psychologists defend human modes of thought by insisting that they must at least have worked well in the environment of evolutionary adaptation, even if they break down in modern environments. This shift of evaluative context, from the modern environment to the evolutionary one, would not be necessary if our modes of thought worked equally well in both, and so implicitly concedes that our biologically natural modes of thought do not work optimally in a wide range of modern situations.

4.2 Ecological rationality, environments and goals

A similar objection may be levelled against the proponents of ecological rationality. Can our ability to reason be identified with an adaptive toolbox composed by fast and frugal
heuristics, modeled together by evolution in order to solve a wide range of relevant problems? Although Gigerenzer’s ecological characterization of rationality seem to be compelling at first sight, Stanovich and West (2003: 206-209) have pointed out that there are strong ambiguities and inconsistencies in his works with regard to whether ecological rationality is to be considered optimization to the environment of evolutionary adaptation or to the contemporary environment. In my view, this ambiguity leads to a more general issue about the characterization of the goal at which, according to the ecological rationality supporters, good reasoning should aim in actual circumstances. Let us consider an interesting example of inconsistency that Stanovich and West have found in the book *Simple Heuristics that Make Us Smart* (1999). In the final chapter of the book, speaking of how the usefulness of heuristics should be assessed, Todd and Gigerenzer (1999: 364; as reported by Stanovich & West 2003: 206) maintain that ecological rationality depends on fast and frugal heuristics that promote “an organism’s adaptive goals in the physical or social environment”. However, Todd and Gigerenzer make no attempt to explain whether they are referring to evolutionary goals (in which case ecological rationality aims at optimizing people’s inclusive fitness in the environment of evolutionary adaptation) or people’s personal goals that are relevant or adaptive in the current social or natural environment. As Stanovich and West state (2003: 206), this ambiguity makes “the term ‘ecological rationality’ difficult to pin down and hence to evaluate”. As regards to the question of rationality assessment, Gigerenzer and his collaborators seem to change the meanings of ecological rationality and its related criteria, passing from an evolutionary to an ecologically adapted notion of rationality, adopting whichever characterization is most convenient for the assessment being made. Sometimes they refer to an adaptive function of rationality but in other occasions they foreground the role of rationality at the personal level, and these two functions of rationality do not always coincide (Stanovich & West
Suppose that someone proposes you a choice between eating a chocolate or a carrot. According to the evolutionary standpoint, an unconscious instinct probably makes you like the foods that have more fat or sugar, and thereby you will have a strong tendency to prefer the chocolate to the carrot. As is noted by Over (2004: 11), “this instinct will be beneficial under primitive conditions, when maximum calorie intake was necessary for reproductive success, but it may damage health in technologically advanced society”. Thus, Over invites to conclude that we cannot say that making this choice now is rational only because “the cognitive structure that brings it about facilitated the goal of reproductive success for our ancestors early in human (and pre-human) evolutionary history” (Over 2004: 11). It appears that Gigerenzer and his collaborators want to ground their account of ecological rationality on evolutionary considerations, without accepting some of their most questionable consequences (Stanovich & West 2003: 208).

Evolutionary arguments as well as ecological arguments for the rationality of human cognition seem to set aside the real conflicts between goals linked, directly or indirectly, to inclusive fitness and “personal” goals, which do not always coincide (Stanovich & West 2003: 172). What it is needed is a distinction between the actual range of goals people aim at and those which come from the evolution of human species. But a picture of human rationality based only on evolutionary considerations does not seem to be able to provide this distinction. Dual rationality theories can be seen as an attempt to reply to such a question.

5. Dual Rationality Theories

Both evolutionary and ecological approaches reject the classical normative standards of rationality. This is however a questionable move: as a matter of fact, empirical research has
demonstrated that there are people (being a minority among the experimental subjects) who are able to comply with the normative principles of rationality across a wide range of reasoning problems.\textsuperscript{7} If so, it is wrong to argue, as evolutionary psychologists do, that the human mind operates exclusively on the basis of domain-specific cognitive mechanisms adapted to their proper domains. But how can we explain such differences among subjects’ reasoning performances? In tackling this question, some cognitive psychologists and philosophers of mind have developed a family of dual-process theories about the cognitive mechanisms underlying reasoning and decision making. While differing as to their details, these theories claim that the human mind possesses two quite different systems of thinking and therefore assume two correspondingly different conceptions of rationality. The proponents of these theories hold that insofar as we recognize that the human mind is composed by two systems working on the basis of two different kinds of rationality, it will be possible to improve our understanding of what rationality amounts to.

\textbf{5.1 The two cognitive systems}

Dual-process theories, which began with the works of Evans and Wason in the seventies (Wason \& Evans 1975), have been developed recently by various researchers in different ways. Most prominently, psychologists and philosophers have applied this approach to deductive reasoning (see Evans 2003; 2006; Evans \& Over 1996; Stanovich, 1999, 2004), inductive reasoning (see Kahneman \& Frederick 2002; Sloman 1996) and decision making (see, e.g., Reyna, 2004).\textsuperscript{8} While differing as to details, these theories agree on a set of

\textsuperscript{7} Stanovich and West have discovered a strong correlation between success in reasoning tasks and good scores on tests of intelligence (such as the Scholastic Aptitude Test (SAT)), which they interpret as indicating that people who have higher scores on tests of intelligence tend to give correct responses to the question of classical reasoning tasks as intended by the experimenter (see, e.g., Stanovich 1999; Stanovich \& West 2000).

\textsuperscript{8} For recent developments of dual-process theories see Evans \& Frankish 2008.
distinctive characteristics which are attributed to the two systems. The so-called System 1, which Stanovich calls also TASS (“The Autonomous Set of Systems”), consists of a collection of different sub-systems that are characterized as fast, unconscious, automatic, operating in parallel together, and requiring relatively little cognitive effort (Stanovich 2004: 37-44). Clearly, the domain-specific heuristics and modules, proposed respectively by evolutionary psychologists and Gerd Gigerenzer, are included in this system, which is taken to be universal and shared with other animal species. While System 1 is automatic and effortless when processing certain types of information, System 2 is more “reflective” and can be activated when needed. This system is slow at processing information, rule-based, serial in nature; it operates consciously and requires significantly more cognitive resource (Stanovich 2004: 44-47). This is the kind of system that has been classically taken to be human reason or rationality. Dual-processes theorists hold that in contrast with System 1 this system has evolved more recently and is an exclusively human property. While there has been a general agreement on the multiplicity of System 1, there is no agreement about whether System 2 is one system or includes sub-systems (Stanovich 2004: 35-36). Keith Stanovich, among others, has argued that System 2 must be a unique central processing system. In his view, as opposed to the processes of System 1 that respond automatically to specific inputs, System 2 processing “[...] allows us to sustain the powerful context-free mechanisms of logical thought, inference, abstraction, planning, decision-making, and cognitive control” (Stanovich 2004: 47). Similarly, Jonathan Evans and David Over (1999) hold that the function of System 2 is to make possible hypothetical thinking, which requires an ability to represent possible states of the world. This ability is fundamental for reasoning and decision making. For example,

9 Many different labels have been used to refer to the two systems depending on what dual-process theorists want to highlight in their theories (see Stanovich 1999: 145). Following Stanovich’s terminology, I will call them System 1 and System 2 (Stanovich 1999: 144-148).
deductive reasoning is hypothetical when its premises are not actual beliefs, but rather assumptions or suppositions. […] Consequential decision making consists of forecasting a number of possible future world states and representing the possible actions available. […] Scientific thinking is itself hypothetical when entertaining hypotheses about the way the world might be and deducing their consequences for making predictions. (Evans & Over 1999: 764)

5.2 Two kinds of rationality

Since the dual process theory is itself a psychological theory, its bearing upon the question of what rationality amounts to remains an open question. However, several researchers within this trend of research have offered suggestions about the implications that the theory could have for the rationality debate (Evans & Over 1996; Stanovich & West 2000; 2003). According to them, if the human mind consists of two sub-systems, in order to assess each of these, we are in need of two different accounts of rationality. In particular, while Keith Stanovich and Richard West (1999) distinguish two senses of rationality that are applicable to procedures that work respectively at the subpersonal and personal levels, Jonathan Evans and David Over (1996) characterize the distinction in terms of the contrast between unconscious heuristic and conscious rule-following processes.

According to Stanovich and West (2000; 2003), System 1 is strictly linked with adaptive rationality, according to which to be rational is to reason so as to maximize inclusive fitness. So, System 1 consists of cognitive mechanisms that are highly successful at making decisions which contribute, directly or indirectly, to the promotion of inclusive fitness. Opposite to that, System 2 includes cognitive mechanisms that operate in accordance with classical normative principles of rationality. According to this approach, we have two senses of rationality, that is, evolutionary and normative rationality. Stanovich and West
hold that what have been considered as evidence of human irrationality in classical studies on human reasoning should be considered as evidence of irrationality in the latter sense but not in the former. On the one hand, in most cases, either of these two kinds of rationality requires the same normatively appropriate response: in such cases, reasoning that is rational according to the classical normative standards of rationality is rational according to evolutionary rationality as well. On the other hand, however, insofar as there is a certain amount of reasoning problems in which the two kinds of rationality do not prescribe the same response and some reasoning performances conform to one response and other performances to the other response, that should demonstrate that our cognitive system includes at least two sub-systems which interact and override each other. As Stanovich and West characterize this distinction, “evolutionary” and “normative” rationality may be viewed as “terms for characterizing optimization procedures operating at the subpersonal and personal levels” (Stanovich & West 2000: 661); and “optimization” here is very naturally read as suggesting a broadly consequentialist conception of rationality.

Even if Evans and Over’s proposal seems to be very similar to that of Stanovich and West, they provide a different characterizations of the two kinds of rationality:

\[ \text{Rationality}_1: \] Thinking, speaking, reasoning, making a decision, or acting in a way that is generally reliable and efficient for achieving one's goal.

\[ \text{Rationality}_2: \] Thinking, speaking, reasoning, making a decision, or acting when one has a reason for what one does sanctioned by a normative theory. (Evans & Over 1996: 1)

They also argue that “people are largely rational in the sense of achieving their goals (rationality$_1$) but have only a limited ability to reason or act for good reasons sanctioned by a normative theory (rationality$_2$)” (Evans & Over 1997: 1). According to Evans and Over, people can achieve goals without following any normative principle of rationality, i.e.,
without having a justification for their reasoning, decision or judgment. But it does not follow from this that the goals mentioned in the characterization of rationality₁ amount to inclusive fitness. System 1 processing can help to achieve very different kinds of goals in actual circumstances, namely not only in the proper domains for which cognitive mechanisms have evolved. In other words, it seems to me that Evans and Over’s characterization of rationality₁ is very close to a consequentialist conception, and is not necessarily centred on the notion of inclusive fitness. So, in contrast with Stanovich and West’s proposal, the distinction made by Evans and Over suggests that we should distinguish between two different kinds of cognitive mechanisms which, since they have different and sometimes competing goals, require two different conceptions of rationality. While rational₁ cognitive mechanisms are characterized in consequentialist (and not only evolutionary) terms, rational₂ cognitive mechanisms are characterized according to a deontological conception of rationality.

5.3 The implications of dual rationality for the evaluative question

It is generally assumed by dual-process theorists that System 1 and System 2 differ in many respects. But then one immediate challenge for dual-process theory concerns the relationships between the two systems. How do they interact with one another? In particular, how can System 2 override cognitive mechanisms belonging to System 1 when they are triggered by the same input?

The question of how the two systems interact with one another is the Achilles’ heel of any dual process theory. For example, Evans and Over give us no detailed or satisfactory account of that. This seems to be a missing detail which Evans himself recognizes. As he writes, “an important challenge is to develop models to show how such two distinct systems interact in one brain and to consider specifically how the conflict and competition
between the two systems might be resolved in the control of behaviour” (Evans 2003: 458).

In particular, what is needed is an account of how the two systems interact together. This account may allow the corresponding dual rationality theory to be used to assess human reasoning. On the one hand, given the consequentialist understanding of rationality it should be easy to understand when System 2 overrides System 1, that is, when doing so helps us to reach our goals in the current context. On the other hand, however, as is noted by Over (2004: 11), “it can be difficult, and sometimes impossible, in practice to decide when to override”. Both System 1 and System 2 are the product of evolution, but only System 2 permits us to pursue goals that are not directly linked to reproductive success. While System 1 is governed by domain-specific and automatic cognitive mechanisms, System 2 permits to figure out how to attain personal goals. It is clear that not all people have reproductive success as one of their goals in System 2, although that goal belongs to them as a goal of System 1 (Over 2004: 10-11). As a result, the two systems are aimed at different and sometimes competing domains of goals. This shows that we cannot characterize rationality simply in terms of inclusive fitness or adaptiveness, or simply replace the classical standards of rationality by evolutionary or ecological ones.

6. Concluding remarks

A good way to ground normative claims about rationality is by appealing to the outcomes that good reasoning should lead to. Consider two of the most important implications that the adoption of a consequentialist approach has for the rationality question. First of all, a consequentialist conception assumes that the norms of rationality are both psychologically dependent and robustly context-dependent. Indeed, contrary to the idea that norms of rationality are universal, consequentialism implies that what good reasoning is may vary
across contexts. Normative standards must be relativized to goals and environments: a reasoning strategy that leads reliably to a good outcome in a given context is, therefore, a normatively appropriate way of achieving that goal in that context, but not necessarily in other contexts. Secondly, according to the consequentialist approach, the normative and the descriptive question (see Chapter 1, Section 1.1) are inextricably interrelated because only a psychological theory of reasoning can identify a collection of reasoning strategies out of which to select the most effective strategies in a given context.

Within this theoretical framework, heuristics, such as those characterized by Kahneman and Tversky, can be justified consequentially because they help us reach two goals: an outcome which the reasoner values and the sparing of time and energy. Inasmuch as successful outcomes are characterized in terms of their conduciveness to inclusive fitness, then it follows that reasoning performances are rational to the extent to which they contribute to increase reproduction and inclusive fitness. As seen above, this standpoint has been supported by two different but related traditions: evolutionary psychology and ecological rationality. Although their proponents have correctly re-evaluated the rationality of heuristic processes, their analyses are based on a misinterpretation of human rationality. As is pointed out by Stanovich and West (2003: 172), “what these theorists have missed (or failed to sufficiently emphasize) is that definitions of rationality must coincide with the level of the entity whose optimization is at issue”. That calls for a distinction between adaptive rationality and “personal” rationality and the goals which are connected to them. Indeed, although evolutionary psychologists acknowledge that the conditions in the environment of evolutionary adaptation do not always match those of our current world, they do not take into consideration the consequences of not clearly distinguishing between goals linked, directly or indirectly, to inclusive fitness and personal goals.
To conclude, consider again the expression “organism’s adaptive goals” which is so often used by both evolutionary psychologists and Gerd Gigerenzer (see Chapter 3, Section 4.2). As is noted by Stanovich and West (2003: 206-207), it is possible to attribute two different meanings to this expression. On the one hand, as evolutionary psychologists hold, the word “adaptive” suggests that we are specifically talking about evolutionary/adaptive goals. On the other hand, we can focus on the word “organism” and view the concepts of rationality proposed by evolutionary psychologists and Gigerenzer as closer to that of consequentialist personal rationality. If “adaptive” goals are taken to refer to the environment in which a reasoner is situated, reasoning performances ought to be assessed in terms of their success in achieving the goals people aim at, given the constraints imposed by their current environment (and not by referring to the environment of evolutionary adaptation). In the next chapter, I will focus on this second interpretation of the expression “organism’s adaptive goals”. In particular, I will explore a consequentialist approach to rationality assessment that is not strictly dependent on the characterization of rationality in terms of inclusive fitness or evolutionary adaptiveness.
Chapter 4 – Consequentialism and Pragmatism

Introduction

In the previous chapters, I have examined two different types of accounts of rationality, respectively committed to a deontological and a consequentialist conception of rationality, and argued that neither is adequate to provide an appropriate normative framework for assessing human reasoning. However, with regard to the consequentialist conception of rationality, I have criticized one of its forms in particular, i.e., the one based on evolutionary considerations.

In this chapter, I explore two consequentialist approaches to rationality assessment that are not strictly dependent on a characterization of rationality in terms of evolutionary adaptiveness but are inspired in different ways by pragmatism, i.e., Stephen Stich’s Epistemic Pragmatism and Michael Bishop and John Trout’s Strategic Reliabilism. In both these approaches, which presuppose consequentialist epistemological frameworks, experimental evidence on human reasoning is taken into consideration in order to argue that humans employ a multiplicity of reasoning strategies and that, in order to comparatively assess the epistemic quality of these and decide which normative standards to adopt, several cognitive and situational constraints ought to be taken into account. Those constraints are then characterized as determined by pragmatist considerations such as the cognitive resources available to reasoners and the value or significance they attach to the goal they pursue.

As to these two consequentialist accounts, in the last part of the chapter, I criticize their tendency to neglect the distinction between cognitive and situational constraints by
grounding their analyses exclusively on the kind of pragmatist considerations mentioned above. That leads me to introduce the notion of situational context, which I characterize as complementary but distinct from the cognitive context assumed by pragmatists such as Stephen Stich. Finally, with regard to what has been called the evaluative question, I point out at the vacuity of any consequentialist approach not taking into serious consideration the role of situational constraints in human reasoning.

1. From the standard analytic approach to epistemology to epistemic consequentialism

Since both the pragmatist approaches to rationality I discuss in this chapter presuppose consequentialist epistemological frameworks, I start by introducing the consequentialist approach to epistemology by differentiating it from the traditional way pursued by most contemporary analytic epistemologists.

1.1 Epistemic consequentialism: an alternative way to approach epistemology

Roughly speaking, the traditional task of epistemology is to define what knowledge amounts to and what makes it possible. According to the predominant view in epistemology, knowledge is conceived as a special kind of true belief, that is, justified true belief. While belief and truth are not really epistemological concepts (they are a psychological and respectively a semantic one), classical epistemological enquiry is centred on the notion of justification. Epistemology should aim to provide an appropriate analysis of the concept of justification and work out the conditions that beliefs must satisfy in order

---

1 The traditional analysis of knowledge has its roots in Plato’s *Theetetus* where knowledge is characterized as “true belief with a reason”. In its general form this analysis may still be found in most contemporary epistemological literature and, in particular, in those belonging to the analytic tradition. The resulting definition is stated as follows: $S$ knows that $p$ if and only if (i) $S$ believes that $p$ (ii) $p$ is true and (iii) $S$ is justified in believing that $p$. The most famous counter-example to the traditional analysis of knowledge has been proposed by Edmund Gettier (1963) (for a recent overview, see Lycan 2006).
to count as justified true beliefs, that is, knowledge. Since the notion of justification is assumed to be normative, epistemological theories are evaluative theories. Their normative criteria are applied to belief tokens, in order to establish whether they are justified or not. In the last twenty years or so, however, a new approach to epistemology has emerged as a rival to the standard analytic ways of doing epistemology. According to epistemic consequentialists, the main aim of an epistemological enquiry is to determine how to properly assess and improve reasoning strategies of belief formation and revision by attending to their consequences (see, e.g., Bishop & Trout 2005; Goldman 1978; Stich 1990). Insofar as a consequentialist approach to epistemology aims to develop a way to evaluate how different people reason, it will not focus on the assessment of belief tokens, but of reasoning strategies. So, in these approaches the notion of justification, which is considered as a property of belief tokens, plays no role. The epistemological enquiry shifts to how people reason, rather than the conditions for knowledge.

Within this new approach to epistemology, an increasing interest has been raised by the status of the epistemic subject and by her psychological and social conditions. Indeed, epistemic consequentialism is usually characterized as a naturalistic approach to epistemology as well as to the question of human rationality. The label “naturalistic epistemology” does not refer to a well-defined epistemological account; rather, it covers a number of very different and competing naturalistic approaches to epistemology. In the last decades there has been a lot of discussion about the appropriateness of naturalistic approaches in epistemological theorizing. Contrary to the so-called replacement thesis (Kornblith 1994b: 4), namely the thesis that epistemology should be set aside in favour of psychology, epistemic consequentialism assumes a moderate position about the naturalization of epistemology. Its supporters do not aim to replace epistemology with

---

2 In the analytic tradition, these analyses are usually grounded on considered judgments about our ordinary concept of knowledge and justification. For a critical discussion about this fundamental methodological question in epistemology see, e.g., Weinberg et al. (2001), Nichols et al. (2003), Bishop & Trout (2005).
psychology (Quine 1969b), because, according to them, epistemology has to do with concerns about epistemic and rational norms rather than with the description of epistemic performances. In particular, epistemic consequentialism provides a normative framework for comparatively assessing the epistemic quality of a given reasoning strategy in terms of its conduciveness to a given epistemic goal. So, this approach makes room for a more substantial notion of epistemic goodness and rationality that takes into account the nature and value of the goals pursued by the epistemic subjects. As Stich (1993: 5) has pointed out, that requires to “[…] determine which goal or goals are of interest for the assessment at hand”: a step which, according to Stich himself, is “fundamentally normative”. Indeed, any empirical enquiry will not explain what people’s goals should be.

1.2 Epistemic goals: truth and beyond

According to consequentialism, epistemologists must determine what the epistemic goals of reasoning strategies ought to be and what reasoning strategies will best lead to those goals given the reasoners’ cognitive and environmental constraints. However, whatever goal consequentialists opt for, they should explain why it matters from an epistemic point of view.

Looking at the epistemological literature, many in the naturalistic tradition have maintained that truth is the most fundamental goal of our epistemic activity (see, e.g., Goldman 1986; Kornblith 2002). Leaving aside his former claim that epistemology should become part of psychology, for example, Quine (1986: 664-665) stated that naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in a more cautiously epistemological term, prediction […]. There is no question here of
ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed.

A well-known form of this sort of epistemic consequentialism is Goldman’s Reliabilism (Goldman 1986; 1999). According to him, good reasoning strategies should be aimed at producing true beliefs, and their outcomes can be taken to be justified if they are generated by reliable psychological processes that generally lead to true beliefs. Accordingly, reasoning strategies are to be assessed by their success in leading to true beliefs. But, as Hilary Kornblith (2002: 123) observes, “how is it that truth acquires this status as our goal and confers normative force to the recommendations to pursue certain reasoning strategies of belief formation and revision, namely, those which are conducive to achieving it?”. Why does truth matter? To these questions, a first answer is that truth has an intrinsic value. That means that holding true beliefs is intrinsically valuable. A second possible answer is that, although an epistemic subject aims at having true beliefs, she does so because having true beliefs may be useful to attain other more valuable goals: truth as an epistemic goal has merely an instrumental value. According to this second view, having true beliefs is valuable because true beliefs help us to achieve the goals we aim at.\(^3\) Alternative to these two truth-centered answers, pragmatists replace truth as the main epistemic goal with more practical purposes. In particular, Stephen Stich (1990) argues that truth should be set aside in favour of other goals, such as the totality of goals people value. According to him, an epistemologist should consider the consequences of using this or that reasoning strategy with respect to their conduciveness to achieving such or such pragmatic goal.

\(^3\) It is implicitly assumed that true beliefs are more conducive to valuable practical consequences than false beliefs are.
2. Stich’s Epistemic Pragmatism

In this section, I begin by presenting Stephen Stich’s Epistemic Pragmatism, then proceed to focus on the implications that the adoption of this view might have for what we have called the evaluative question, and finally criticize the two main theses of Epistemic Pragmatism, that is, the irrelevance of truth as an epistemic goal and the relativistic value-based approach to reasoning performance assessment.

2.1 Descriptive and normative cognitive pluralism

In The Fragmentation of Reason (1990), Stich argues for what he calls “cognitive pluralism”. In his view, cognitive pluralism is divided into two theses, that is, descriptive and normative cognitive pluralism. According to descriptive cognitive pluralism, people differ significantly in their ways of reasoning, and forming and revising beliefs. Contrary to that, a supporter of descriptive cognitive monism would hold that, if there are differences in how people reason, they will be not significant, and thereby she would conclude that all people reason in fundamentally the same way (Stich 1990: 13). While descriptive cognitive pluralism and descriptive cognitive monism are based on empirical considerations (they do not have any normative import), normative cognitive pluralism is about the reasoning strategies people ought to use. In particular, this thesis holds that while people use a variety of reasoning strategies which significantly differ from each other, they may all be normatively appropriate. In opposition to that, a supporter of normative cognitive monism would hold that there is only one normatively appropriate way of reasoning, regardless of whether different people use different and sometimes competing reasoning strategies. According to the normative monist, you can always find universal criteria which distinguish between correct and faulty ways of reasoning. A clear example of such a

4 Similarly, ethicists distinguish between descriptive ethical relativism and normative ethical relativism.
position is provided by the supporters of what Stein has called the *standard picture of rationality* (see Chapter 1, Section 2.1). So, coming back to the general idea underlying cognitive pluralism, Stich maintains not only that (i) people reason in different ways but also that (ii) there is no single, universal normative standard for assessing which way of reasoning is better than another.

To begin with (i), what is the evidence for descriptive cognitive pluralism? When Stich wrote *The Fragmentation of Reason*, he held that the main empirical evidence for descriptive cognitive pluralism came from studies of human reasoning made by cognitive psychologists such as Peter Wason, Daniel Kahneman and Amos Tversky. According to Stich, this experimental research shows that different people within the same culture use different reasoning strategies or heuristics and tend to go on even when they were explained that their reasoning strategies are normatively inappropriate for a given task. This is a controversial claim, however. Nenad Miščević, for example, has noted that

> there is no variety of wrong answers given at the selection or conjunction tasks, and this uniformity is a testimony to the importance of the experimental paradigm itself. The tests have been performed on people of very different degrees of sophistication and age, but the biases seem to be uniform; they do not, at least *prima facie*, support any kind of descriptive pluralism. (Miščević 1996: 28)

According to Miščević, the results of the classical research on human reasoning can be interpreted as indicating that almost all people possess and use the same heuristics in order to solve reasoning problems and so, he concludes, such data might support, contrary to Stich’s claim, descriptive cognitive monism. However, in the classical research on human reasoning most, if not all, experimental subjects were Western people. At that time, no
systematic research concerning people of different cultures was done on reasoning strategies that might really support descriptive cognitive pluralism.

Since the late nineties, cognitive psychologists have produced new evidence which seems to support Stich’s claim. In particular, the social psychologist Richard Nisbett and his collaborators have conducted a number of psychological experiments to test whether “Western” and “East Asian” people think and reason differently when faced with the same cognitive task (see, e.g., Nisbett et al. 2001; Nisbett 2003). Their results show significant differences among cognitive (including reasoning) strategies used by Westerners and East Asians. According to their proponents, these studies question the idea that all people share a basic core of cognitive strategies regardless of their own culture and education.

In particular, Nisbett and his collaborators characterize the cognitive strategies used by Westerners as being more analytic. Such a way of reasoning involves

[…] detachment of the object from its context, a tendency to focus on attributes of the object to assign it to categories, and a preference for using rules about the categories to explain and predict the object’s behaviour. Inferences rest in part on the practice of decontextualizing structure from content, the use of formal logic, and avoidance of contradiction. (Nisbett et al. 2001: 293)

Instead, cognitive strategies employed by East Asians are characterized as being more holistic. Their ways of reasoning involve

[…] an orientation to the context or field as a whole, including attention to relationships between a focal object and the field, and a preference for explaining and predicting events on the basis of such relationships. Holistic approaches rely on experience-based knowledge rather than on abstract logic and are dialectical, meaning that there is an emphasis on change, a recognition of contradiction and of
the need for multiple perspectives, and a search for the “Middle Way” between opposing propositions. (Nisbett et al. 2001: 293)

Consider two strands of experimental research developed by Nisbett, Norenzayan and their collaborators.

**Deductive reasoning and logic.** Starting from the results obtained from a series of empirical studies on subjects’ evaluations of deductive arguments, Nisbett and his collaborators have argued that Westerners are more logical than East Asians. In particular, according to them, East Asians “might be expected to rely more on prior beliefs and experience-based strategies when evaluating the convincingness of deductive arguments than do Westerners” (Nisbett et al. 2001: 296). In contrast, Westerners should “be more capable of ignoring prior beliefs and setting aside experience in favour of reasoning based on logical rules” (Nisbett et al. 2001: 296). In the book *The Geography of Thought* (2003), Nisbett presents two experiments in which East Asian and Western subjects are given a series of arguments and asked to evaluate whether they are logically valid or how convincing they are.

In the first experiment, Norenzayan and his collaborators (2002: 668-672) asked Korean and American subjects to judge the convincingness of arguments such as the following

1. All birds have ulnar arteries;
   Therefore all eagles have ulnar arteries.

2. All birds have ulnar arteries;
   Therefore all penguins have ulnar arteries.\(^5\)

---

\(^5\) *Ulnar artery* is a “blank” property, used by the experimenters so that real-world knowledge cannot affect the assessment of the deductive argument (Nisbett 2003: 168).
By using such arguments, Norenzayan and his collaborators would like to examine how their subjects “[...] ‘project’ properties – ‘ulnar arteries’ in the above example – from superordinate categories (birds) to subordinate categories (eagles, penguins)” (Nisbett 2003: 168). It is noteworthy that (1.) and (2.) have the same premises but their conclusions differ with respect to the typicality of the target bird: penguins are less typical birds than eagles. In order to reason about these arguments, two different strategies may be employed by subjects. Those who follow the principles of logic would acknowledge the existence of hidden premises in each argument, that is, “All eagles are birds” and “All penguins are birds”. As Norenzayan and his collaborators (2002: 668) explain, “once these hidden premises are exploited, the argument becomes a standard valid deductive argument. Armed with this knowledge, participants should be equally convinced by the typical and atypical argument”. Most Korean subjects, instead, tend to consider (1.) more convincing than (2.), because “prior experience makes them more comfortable with regarding eagles than regarding penguins as birds” (Nisbett 2003: 168-69). As expected by Norenzayan and his collaborators, Korean subjects’ evaluations are influenced by the typicality of the target bird as they find more convincing the typical arguments than the atypical ones, American subjects were convinced by both kinds of arguments.⁶

In another experimental study (Norenzayan et al. 2002: 672-678), subjects were asked to judge the logical validity of a series of arguments that were either deductively valid or invalid and whose conclusions were either believable or non-believable (see Arguments A-B). At the end of the task, the same argument structures were presented in an abstract version so as to evaluate their logical validity (see Argument C).

⁶ As Nisbett and his collaborators (2001: 301) point out, “when an experimental manipulation was introduced that increased the salience of the typicality information, all three groups showed the typicality effect to the same extent".
Argument A

Premise 1: No police dogs are old.
Premise 2: Some highly trained dogs are old.
Conclusion: Some highly trained dogs are not police dogs.

Argument B

Premise 1: All things that are made from plants are good for health.
Premise 2: Cigarettes are things that are made from plants.
Conclusion: Cigarettes are good for health.

Argument C

Premise 1: No A are B.
Premise 2: Some C are B.
Conclusion: Some C are not A.

While both Argument A and argument B are meaningful, the conclusion of the former is plausible and that of the latter is not plausible. Argument C is, instead, abstract. It is noteworthy that all of them are logically valid. The task of assessing their logical validity was presented to American and Korean college students. Both groups had good performances on the abstract syllogisms, but as regards the first set of arguments (Arguments A and B), Korean subjects were more prone to belief bias than American subjects. As Nisbett, Peng, Choi and Norenzayan (2001: 301) point out, when Korean subjects were given arguments that had implausible conclusions, they displayed a stronger tendency to classify them as valid: “the results indicate that when logical structure conflicts with everyday belief, American students are more willing to set aside empirical belief in favour of logic than are Korean students”.

133
Attitudes to contradiction. According to Nisbett (2003: 173-185), when faced with seemingly contradictory statements, East Asians may accept both. In an experimental study conducted by Peng and Nisbett (1999: 748-749), the experimenters presented American and Chinese subjects either with one statement or with two statements that, albeit not straightforwardly contradictory, were very different and apparently unlikely to be both true. These statements were presented as if they were part of a social science research. For example, statement A was:

A survey found that older inmates are more likely to be ones who are serving long sentences because they have committed severely violent crimes. The authors concluded that they should be held in prison even in the case of a prison population crisis. (Peng & Nisbett 1999: 754)

Its counterpart (statement B) was:

A report on the prison overcrowding issue suggests that older inmates are less likely to commit new crimes. Therefore, if there is a prison population crisis, they should be released first. (Peng & Nisbett 1999: 754)

In the experiment, subjects were given one of these statements (condition “A or B”) or both (condition “A and B”) and then asked to rate their plausibility. In condition “A or B”, both Chinese and American participants found A more plausible than B, while in the “A and B” condition, “Americans judged the plausibility of the more plausible proposition as greater than did Americans who read only the more plausible assertion by itself” (Nisbett et al. 2001: 302). This means that American subjects judged a contradicted statement to be more plausible than the same statement considered in isolation. In contrast, Chinese subjects in
the “A and B” condition judged the two conflicting statement to be equally plausible, as if they wanted to find merit in both. As is pointed out by Nisbett and his collaborators (2001: 302), Chinese subjects deemed the less plausible statement to be more plausible when it had been contradicted than when it had not. Such a judgement is again normatively problematic, but very different from that of the American subjects.

So, on the basis of these experimental studies, Nisbett and his colleagues hold that “literally different cognitive processes are often invoked by East Asians and Westerners dealing with the same problem” (Nisbett et al. 2001: 305). Moreover, according to them, such differences between ways of thinking can be used as evidence that people not only use very different cognitive strategies, but they also differ in their beliefs about how the world is. It is easy to draw from these studies the conclusion that the differences in cognitive strategies between Western and East Asian subjects are relative to their cultures and to the social systems to which they belong, hence that their experimental results support Stich’s descriptive cognitive pluralism. As said above, descriptive cognitive pluralism is, however, a descriptive thesis and, as it stands, does not necessarily lead to normative relativism about rationality. The fact that people from different cultures use different reasoning strategies when dealing with the same reasoning problem does not mean that their reasoning strategies are all equally good.

2.2 Normative cognitive pluralism: beyond truth as the main epistemic goal

Why should epistemologists be interested in descriptive cognitive pluralism? According to Stich (1990: 74), the existence of significant differences among people’s cognitive strategies requires the development of an evaluative framework for the comparative

---

7 Although many philosophers and psychologists have raised objections against the works of Nisbett and his colleagues, for the sake of the argument, I will take for granted here both Nisbett and his colleagues’ experimental results and their interpretations (for some of these criticisms, see, e.g., the three commentaries and the reply of Nisbett and Kaiping Peng on American Psychologist (September) 2000: 1064-1068; Engel 2007; Huss 2004; Mun Chan & K.T. Yan 2007).
assessment of those cognitive strategies and their respective epistemic merits. This project begins with a fundamental question: how can epistemologists assess the different ways people reason? Referring to the studies of Nisbett and his collaborators, other similar questions can be raised: what can we say about the normative status of different systems of reasoning strategies such as those demonstrated by Nisbett and his colleagues? Is one of them right and the rest of them wrong from an objective point of view? In supporting his normative cognitive pluralism, Stich wants to show that there is epistemic incommensurability between different ways people reason, that is, there exists no universal criterion for distinguishing between good and faulty ways of reasoning. In consequentialist terms, that means that there is no common goal that might enable us to compare reasoning strategies’ effectiveness across situations. However, in order to defeat Stich’s relativist conclusion, a monist consequentialist might appeal to some common epistemic goal as universal criterion of evaluation, and that might be the goal of having true beliefs. While accepting that people of different cultures and societies reason in significantly different ways, the monist consequentialist holds that all their reasoning strategies aim at truth. So, according to him, if we want to comparatively assess different reasoning strategies with respect to a given epistemic goal, the best candidate will be always the truth. As seen above (see Chapter 4, Section 1.2), these considerations are based on the idea that truth is the fundamental goal, either intrinsically or extrinsically or both, for all cognizers (humans and even animals, regardless of their own culture or society). Against such a monistic epistemological position, Stich has provided two arguments, which show that having true beliefs may not be really valuable either intrinsically or instrumentally (Stich 1990: 101-127).

Consider his first argument, which aims to demonstrate that the notion of truth (and its related interpretation function) is idiosyncratic and culturally bound and, as such, cannot be
used as a universal standard for evaluation. On this view, people should not care whether their beliefs are true rather than other competing plausible (and even counter-intuitive) options such as TRUE*, TRUE**, or TRUE*** etc. The fulcrum of Stich’s argument rests on “the existence of a function that maps certain brain-state tokens (including beliefs and perhaps some others) onto entities that are more naturally thought of in semantic terms, entities like propositions, or content sentences, or specifications of truth-conditions” (Stich 1990: 104). In other words, what Stich also calls “interpretation function” maps certain brain-state tokens onto entities that, he assumes, can be true or false. Thus, for example, such interpretation function maps a brain-state token, such as a belief, onto the proposition “The cat is on the mat”. The interpretation function attributes a content to the belief, that is, that the cat is on the mat. According to Stich’s argument, the belief will be true if and only if the proposition “The cat is on the mat” (to which it is mapped) is true. However, as Stich (1990: 114) points out, “a function is just a mapping, and if the items in one set can be mapped to the items in another set in one way, they can be mapped in many ways”. This means that there might always be an indefinite number of possible interpretation functions, according to which we can map brain-state tokens, such as beliefs, onto propositions. Coming back to the previous example, we could map the belief that the cat is on the mat onto many different propositions, such as “The cat is on the table” or “The cat is on the table in the kitchen”. But which interpretation function is the right one among them? What make it so? According to Stich, in characterizing the “right” interpretation function, analytic philosophers, such as epistemologists and philosophers of language, aim at examining “the judgments of the man or woman in the street about what content sentences or truth conditions get paired with the ordinary beliefs of ordinary folk” (Stich 1990: 105). In the standard analytical approach to epistemology, the people epistemologists refer to belong to a very definite culture and society. So, Stich holds, the interpretation function
sanctioned by their judgments will be very idiosyncratic, and probably differ from that sanctioned by the considered judgments of people belonging to other cultures and societies (see for an empirical support to this thesis Weinberg et al.: 2001). According to this view, there is not only one but many competing interpretation functions and “the fact that we have inherited this idiosyncratic interpretation function rather than some other one is largely a matter of cultural and historical accident” (Stich 1991: 138). As a consequence, Stich holds that there is nothing intrinsically valuable in having beliefs that are mapped on true propositions sanctioned by the idiosyncratic interpretation function because

...those who find intrinsic value in holding true beliefs [...] are accepting unreflectively the interpretation function that our culture [...] has bequeathed to us and letting that function determine their basic epistemic value. In so doing, they are making a profoundly conservative choice; they are letting tradition determine their cognitive values without any attempt at critical evaluation of that tradition. (Stich 1990: 120)

Stich holds that, while traditionalists, who like to be conservative in epistemic matters, may feel their claims reinforced by that argument, most people, once they are led to understand what is involved in intrinsically valuing true belief, will realize that they do not usually do so.

Consider now his second argument against truth as the main epistemic goal. Stich (1990: 121-124) not only argues that having true beliefs is not intrinsically valuable, he also does not find any good reason to assume that holding true beliefs has an instrumental value. According to him, the fact that true beliefs are good at achieving one’s goals does not mean that they are more intrinsically valuable than other competing options, such as TRUE* beliefs, TRUE** beliefs or even false beliefs. So, Stich argues that we should not focus on whether true beliefs that are in certain cases instrumentally valuable are good at achieving
one’s goals, but rather whether true beliefs that are sanctioned by our idiosyncratic interpretation function are more intrinsically valuable than those assumed to be true by other competing interpretation functions. If these alternative options give rise to different advice about what to do in a given situation and so lead to take different courses of action, they might prove to be more instrumentally valuable than our idiosyncratic notion of truth.

Stich makes reference to one example about one’s survival where having true beliefs can be less useful than having false beliefs. Suppose that a man, call him Harry, rightly believes (he has a true belief) that his plane is scheduled to take off at 7:45 a.m. He arrives at the airport just in time, receives the boarding-card at the check-in, boards the plane, but, after taking off, the plane goes down and crashes killing Harry. In such a case, Stich argues, having a not true belief, for example, that the plane is scheduled to take off at 8:15 a.m., would have saved Harry’s life, that is, having a not true belief would have helped him achieve a basic goal, that is, his survival. Stich proposes this example in order to show that having false beliefs sometimes help us to achieve our (fundamental) goals more than having true beliefs, and this is quite clear from the example. Stich concludes his argument by stating that “the instrumental value of true beliefs is far from obvious, and those who think that true beliefs are instrumentally valuable owe us an argument that is not going to be easy to provide” because “it is surely not the case that having true beliefs is always the best doxastic stance in pursuing goals” (Stich 1990: 124).

2.3 Pragmatism and relativism

If truth is not the common epistemic goal of people’s activity, and so it cannot be used as the fundamental criterion to evaluate good reasoning, what can? If systems of reasoning of different cultures are epistemically incommensurable, how can we adjudicate between them? Stich proposes to replace truth, understood as an absolute cognitive value, with an
indefinite multiplicity of values which are relative to people’s preferences and those of the societies which they belong to, and which may be in competition with one another:

if the argument about the value of truth could be sustained, the natural upshot for the normative theory of cognition would be a thoroughgoing pragmatism which holds that all cognitive value is instrumental or pragmatic – that there are no intrinsic, cognitive values. (Stich 1990: 21)

On his pragmatist view, the reasoning strategy to be preferred is the one “that would be most likely to achieve those things that are intrinsically valued by the person whose interests are relevant to the purposes of the evaluation” (Stich 1990: 131). In other words, good reasoning strategies for a reasoner to employ are those that are more conducive to the state of affairs she considers intrinsically valuable, that is, states of affairs such as, according to Stich (1990: 25), those that help people to control nature or that contribute to improve their living conditions. In his view, reasoning strategies should be deemed to be cognitive tools and evaluated consequentially, that is, in terms of their effectiveness in attaining things which people who use them intrinsically value, regardless of their producing true beliefs.

Within this pragmatist framework, it is possible to better understand Stich’s relativistic view. According to him, neither the goals nor the means to achieve those goals will be the same for all reasoners. People aim at different and competing goals depending on what their desires and interests are, and desires and interests usually vary from person to person and from culture to culture. On this view, it seems impossible to find a single normative model for assessing people’s reasoning strategies. So, insofar as Nisbett and his collaborators’ claims are well supported empirically, the only method available to us for assessing people’s reasoning strategies is to use situational or personal standards of rationality.
2.4 Epistemic Pragmatism’s implications for the evaluative question

What implications does the adoption of such a pragmatist view have for what we have called the evaluative question? According to Epistemic Pragmatism, reasoning strategies can only be assessed consequentially by examining how efficiently they are likely to satisfy one’s desires and personal goals, that is, by attending to their consequences. It is noteworthy that, contrary to the deontological conception of rationality, such an account provides a justification of rationality. According to Richard Samuels, Stephen Stich and Luc Faucher (2004: 166), indeed, it explains why people should aim at reasoning in a normatively appropriate way, that is, why it confers normative force to the recommendation to pursue certain reasoning strategies of belief formation and revision rather than others. The explanation is simple and clear: good reasoning is desirable because it helps us achieving what we intrinsically value. According to it, we aim at reasoning well because this is a necessary condition for attaining things which we intrinsically value. It is not, as many epistemologists maintain, our aiming at truth that by itself explains why reasoning in a normatively correct way matters; rather, it is our desire and interest to attain our goals which gives normative force to the adoption of certain reasoning strategies.

After providing Stich’s pragmatist theoretical background, what we now need is to understand how his account can be applied to rationality assessment. Stich observes that when we ask whether subjects are reasoning well, perhaps what we really want to know is whether their cognitive system is at least as good as any feasible alternative, where an alternative is feasible if it can be used by people operating with some appropriate set of constraints. (Stich 1990: 154)

First of all, when assessing reasoning strategies, we are comparing one reasoning strategy to many other competitors. In this sense, no reasoning strategy is normatively appropriate
or inappropriate in any absolute sense. In evaluating a given reasoning strategy, we should compare it to alternatives that are feasible (in contrast with any other logically possible alternative reasoning strategies). On this view, before any negative evaluation about reasoning strategy is being made, one must be sure there is an alternative that “is both pragmatically superior and feasible” (Stich 1990: 156). Thus, several cognitive and situational constraints ought to be taken into consideration when deciding which normative standards of rationality to adopt. But which constraints should we count as appropriate? Stich states that

in deciding which constraints are relevant, or which alternative cognitive systems we will count as feasible, we must look to our purposes in asking the question. Or, as William James might put it, we must ask what the “cash value” of the question is – what actions might we take as the result of one answer or another. (Stich 1990: 155)

What reasoning strategies a reasoner should adopt depends upon her desires, goals, and preferences in various ways and whether one reasoning strategy is appropriate in order to successfully solve a reasoning task will always depend in part on what questions she wants to answer. Here it is the fulcrum of the consequentialist evaluative framework for assessing human reasoning and rationality which comes from Stich’s Epistemic Pragmatism:

the pragmatic assessment of a cognitive system will be sensitive to both the value and the circumstances of the people using it. Thus it may well turns out that one cognitive system is pragmatically better than a second for me while the second is pragmatically better than the first for someone else. (Stich 1990: 25)
Any assessment of a reasoning strategy should be sensitive to people’s values and the circumstances in which they use it (Samuels et al. 2004: 167). Starting from these considerations, two fundamental types of constraints can be picked out: (i) good reasoning is characterized in terms of its conduciveness to achieving one’s desires and goals; (ii) reasoning strategies’ evaluations should be relativized to specific ranges of contexts. As to (i), we need first identify the goals which people value. Once these are specified, we need to determine what reasoning strategies best serve these goals. With regard to (ii), we need specify the kind of cognitive and situational constraints relative to which reasoning strategies’ assessments should be made.

2.5 Against Epistemic Pragmatism

Since Stich’s consequentialist approach to rationality assessment precludes the possibility to draw stable and clear conclusions about the epistemic quality or rationality of a given reasoning strategy, I maintain that it is not a satisfying response to what we have called the evaluative question. As to Stich’s Epistemic Pragmatism itself, I find it questionable because, as I will try to show, the two main theses on which it is grounded, that is, (i) the irrelevance of truth as an epistemic goal and (ii) the relativity of any reasoning strategy assessment, either do not hold or at least are not well-supported.

2.5.1 The value of truth

Stich dismisses truth as the central epistemic goal of cognition on the grounds that, once we understand what truth is and compare it to some of its competitors, we should accept that truth is not as valuable as we have previously thought. I would focus here on what is one of the most controversial point in Stich’s argument against truth as the main epistemic goal. He holds that there might be situations in which having beliefs that are not true (whether we
call that false, TRUE*, TRUE**, TRUE*** etc.) is more conducive to the things which we intrinsically value than having true beliefs. As seen above, Stich provides only one example about such kind of situations. However, let us assume that these situations may occur and that there may exist beliefs that are not true (they can be false, TRUE*, TRUE**…, TRUE*** etc.) and through which we can achieve the things we intrinsically value more than relying on true beliefs. Following Stich, we should assume that people who have those beliefs, sanctioned by the appropriate interpretation function, will have a better life in the long run (and maybe also in short-term). This is a logically possible situation. But is it also realistic? My contention is that Stich needs explain what these alternatives to the classical notion of truth are, how they can be characterized, so that we can compare them with truth and thereby conclude which of them are better at achieving our goals. However, if you look at his theory, you can see that he does not give any detail about TRUTH*, TRUTH** and TRUTH***. If it is only logically possible that having not true beliefs is more useful than having true belief in order to best achieve our goals, that is not still enough to change our mind and prompt us to aim at having not true beliefs (true* beliefs, true** beliefs, false beliefs etc.). One thing is to say that these beliefs exist and are (maybe) identifiable; another is to say that people are able to find reliable and feasible strategies to arrive at those types of beliefs. Furthermore, how can we distinguish between these different types of beliefs? The conditional statement held by Stich, according to which if we had these types of beliefs, they would lead us to reliably achieve our goals (even if we do not know we have them), is not still enough to make us change our reasoning strategies by replacing our notion of truth and its related interpretation function with an alternative one. Before changing our ways of reasoning, we need know something more about what kind of beliefs we should aim at and, more prominently, whether our cognitive capacities are enough to figure out what these beliefs are and, in the case, to achieve them. So, it is Stich that has to
show us what such alternatives to truth amount to: the burden of the proof is on him and the supporters of his thesis. It is possible then to conclude that we have no reason to dismiss truth as our main trans-cultural epistemic goal and hence to consider it as a relevant criterion in order to assess the epistemic quality of a given reasoning strategy of belief formation and revision. However, I want to take one point of Stich’s argument: it is true that we do not all agree on how truth should be characterized and sometimes, given the context where we are located, we have different intuitions about the truth value of a sentence.

2.5.2 Pragmatism and value-based assessments

There is another objection that comes quite naturally up against Stich’s Epistemic Pragmatism, namely, that it leads to a radical form of relativism. According to Edward Stein (1996: 242), such an approach, which presupposes what he calls the relativist picture of rationality, assumes that “what counts as rational is indexed to each human being, so what counts as rational is (at least potentially) different from each human being”. Characterized in such a way, Stich’s position leads directly to nihilism, giving up any attempt to distinguish between good and bad reasoning strategies, and so “anything goes”. So, once we accept that there is no external and independent standard against which to assess people’s reasoning, we are posed in a situation where epistemic anarchy rules. However, that is too extreme a characterization of Stich’s proposal. As seen above, he holds that the consequentialist approach provides normative standards, but they are relativized to cognitive and situational constraints and are applied using pragmatist criteria. That is what Stich means when he speaks of reasoning strategies “used by people operating within some appropriate set of constraints”. Even if Stein has missed the point, however, I think that starting from his considerations we can develop a stronger objection against the consequentialist approach to rationality assessment set by Stich’s Epistemic Pragmatism.
This objection begins with two general questions: what is the status of Stich’s consequentialist approach to rationality assessment? Is such an approach universally applicable?

As argued for by Michael Bishop (2009: 120), “Stich is a pluralist about a great many things, but when it comes to normative, evaluative matters, he is a methodological monist. Regardless of the item one is evaluating, the evaluative considerations that arise are the same: what is most likely to bring about those things one intrinsically values?”. The very fact that this question is always relevant may be held to provide a general and universal criterion for evaluating different reasoning strategies. As is suggested by Baghramian (2003: 176), Stich’s consequentialist approach, notwithstanding its professed relativism, may be therefore held not to be completely relativistic. Its invariant pragmatist criterion could be characterized in the following way: to the extent that a reasoning strategy helps people to achieve what they intrinsically value, it counts as rational. Putting it in comparative terms, if one reasoning strategy is more effective than another in a given situation, it is the better. But then, as Baghramian (2003: 176-177) observes,

Stich needs to address the question of what the criteria are for adjudicating between the right and the wrong applications of the pragmatic criterion. Does Stich allow any evaluative question about the state of affairs and activities that people typically value? How do we assess the value of being able to predict or control nature or contributing to an interesting and fulfilling life? He argues that such values may vary across individuals and cultures, but would he allow that some individuals or even societies may be mistaken in their perception and characterization of what counts as an interesting or fulfilling life? If yes, then we will be relying on something more than the situationally and contextually given criterion of evaluation. If, on the other hand, the argument is that such values are to be treated as primitives, and not amenable to external scrutiny, they are deprived of the means
of evaluating different approaches to cognition – and the cognitive nihilism that
Stich was trying to avoid would loom large.

If the consequentialist-pragmatist approach Stich advocates cannot deal with disagreements
as to the ultimate values, the pragmatist criterion for rationality assessment it provides
appears in some sense incomplete and, therefore, flawed.

3. Strategic Reliabilism

As we have seen, Stephen Stich offers a pragmatist alternative to the standard analytic
approach to epistemology. In this section, I examine another epistemological
consequentialist account which, however, acknowledges the instrumental value of truth.
This account has been recently proposed by Michael Bishop and John Trout. According to
their Strategic Reliabilism, epistemic evaluations are framed in terms of the extent to which
reasoning strategies yield a high amount of true beliefs that are significant for the reasoner
without requiring too many cognitive resources. So, Strategic Reliability assumes that
pragmatist considerations like resource limitations and significance play a fundamental role
in assessing the epistemic quality of a given reasoning strategy.

3.1 Strategic Reliabilism as a naturalistic approach to epistemology

Strategic Reliability is a naturalistic approach to epistemology, which provides a
normative framework for both assessing how different people reason and picking out
reasoning strategies with the greatest epistemic excellence (Bishop & Trout 2005; 2008). In
contrast with Stich’s thesis that truth cannot be our main epistemic goal, Bishop and Trout
hold that Quine was right when he held that epistemology is “the technology of truth-
seeking” (Quine 1986: 664; see Chapter 4, Section 1.2). By taking Quine’s claim seriously,
they want to show that it is possible to extract normative claims from empirical findings coming from psychology (Bishop 2009). In their view, however, Quine was mistaken in holding that epistemology should be identified with behaviouristic psychology. In contrast, when they speak of empirical findings coming from psychology, Bishop and Trout refer to a set of empirical inquiries which they group together under the label “Ameliorative Psychology”. In their view, what characterizes all the trends of research grouped under this label is that they are constitutively normative in the sense that they (implicitly or explicitly) provide “ought” claims that are relevant to how people ought to reason, and hence epistemologists “have to uncover and articulate [their] normative, epistemic principles” (Bishop and Trout 2005: 15). Accordingly, in their view, epistemology should turn into a discipline similar to any branch of the philosophy of science such as, for example, the philosophy of physics. Its main aims will be both (i) to make explicit the fundamental concepts employed in these empirical studies, such as that of good reasoning, and (ii) to systematize the epistemic generalizations that guide the prescriptions derived from their empirical data.

3.2 Evidence-based criteria of evaluation

With regard to (i), Bishop and Trout articulate what they call “the basic building blocks” of their normative framework. They preliminarily delimit the domain of reasoning strategies defining them as rules for making predictive judgments, which consist of four basic elements: (a) the cues used to make the prediction; (b) the formula for aggregating these cues; (c) the target of the prediction; (d) the range of objects to which the strategy may appropriately apply (Bishop & Trout 2005: 71-72). Suppose now a subject faces a reasoning problem. Every reasoning problem can be solved through different reasoning

---

8 Ameliorative Psychology includes empirical research in psychology as well as works in statistics and Artificial Intelligence.
strategies grounded on different cues. But, by using one or another reasoning strategy, this subject may attain different outcomes. So, while aiming at the same target, a subject can employ several competing reasoning strategies which can lead to very different outcomes as a consequence of picking out some cues or others. But which are the excellent ones among them? According to Bishop and Trout, the excellence of a reasoning strategy does not amount to its compliance with normative principles coming from logic or standard probability theory. Those theories of reasoning have no special epistemic status. Rather, they provide reasoning strategies which have to be empirically evaluated as any other reasoning strategy. Indeed, what Bishop and Trout (2005: 71) call the epistemic excellence of a reasoning strategy “involves the efficient allocation of cognitive resources to robustly reliable reasoning strategies applied to significant problems”. This definition identifies three qualities that contribute to the excellence of a reasoning strategy: (i) its reliability with respect to a wide range of problems; (ii) its computational tractability; (iii) the significance of the problem which the strategy is applied to (Bishop & Trout 2005: 55). For the moment, consider the first two qualities, which, Bishop and Trout hold (2005: 79-92), can be evaluated within a cost-benefit framework. On the one hand, the epistemic benefits of a reasoning strategy are quantified in terms of its reliability. The reliability of a given reasoning strategy is a function of its success on its expected range of problems, that is, the range of problems to which any of its users would be disposed to apply it. On the other hand, Bishop and Trout characterize the costs in terms of time, that is, the time it takes to implement a reasoning strategy (start-up costs) or to use that strategy to solve a problem (execution costs). According to them, for example, a reasoning strategy based on Bayes’ Theorem should be considered as not particularly excellent because, although it may be very reliable, its start-up and execution costs are too high (i.e., it is very hard to learn and very costly to apply) (Bishop & Trout 2005 133-134).
Bishop and Trout (2005: 11-16) hold that in order to construct excellent reasoning strategies, attention has to be paid to the experimental findings of Ameliorative Psychology. These experimental findings show that simple reasoning strategies, such as statistical prediction rules, regularly lead to better results than those made by human experts in a wide variety of situations, such as, for example, diagnosing psychiatric disorders and predicting success or failure at school or in business.

Take as an example the so-called Goldberg Rule, which is considered very successful in predicting whether a psychiatric patient should be diagnosed as neurotic or psychotic on the basis of her scores on five scales of the Minnesota Multiphasic Personality Inventory (MMPI). Lewis Goldberg (1965; 1968) constructed the following formula

\[ N = (L + Pa + Sc) - (Hy - Pt). \]

Given the MMPI scores of a psychiatric patient, this rule is very simple to use: clinical judges have only to add her scores from three scales (that is, two personality scales, Paranoia and Schizophrenia, and the validity scale L (Lie)), and then subtract her scores from other two personality scales (Hysteria, Psychasthenia). If \( N \) is less than 45, then the patient is diagnosed as neurotic. Otherwise, the patient is diagnosed as psychotic (Bishop & Trout 2005: 14-15). In an experimental study, Lewis Goldberg compared the predictive accuracy of this rule with that of 29 clinical judges who examined the same material (the MMPI scores of a group of psychiatric patients who had been previously diagnosed as either neurotic or psychotic) and tried to made the same kind of diagnosis. The experimental results showed that the rule constructed by Goldberg made more accurate predictions than those made by the clinical judges. While the accuracy of the clinical judges varied from a low of 55% to a high of 67%, the Goldberg Rule led to an accuracy rate of 70% (reported by Bishop & Trout 2005: 14). Generalizing this and other similar
experimental findings, Bishop and Trout (2005: 17-19) hold that insofar people aim at becoming excellent reasoners and so at having at their disposal a set of excellent reasoning strategies to handle with significant problems, they have to consider what reasoning strategies Ameliorative Psychology recommends us to use.

This passage from empirical data to normative advice might seem arbitrary: how is it that empirical evidence confers normative force to the recommendation to pursue certain reasoning strategies rather than others? Bishop and Trout (2005: 20) hold that this passage is granted by what they call the Aristotelian Principle, which says that “in the long run, poor reasoning tends to lead to worse outcomes than good reasoning”. This means that insofar as, given a definite problem, a reasoning strategy based on a statistical prediction rule tends to lead to good outcomes more often than any reasoning strategy based on experts’ considered judgments, people should use it. Referring back to the previous example, as shown by empirical research, using the Goldberg Rule will lead to more accurate diagnostic predictions than relying on judgments made by clinical judges. So, according to Strategic Reliabilism, the Goldberg Rule is the epistemically best way for predicting whether a psychiatric patient is neurotic or psychotic and ceteris paribus clinical judges ought to use it (Bishop & Trout 2005: 15).

3.4 Truth and good outcomes

As any consequentialist approach to epistemology, Strategic Reliabilism requires a characterization of the notion of good outcome. What is it that Bishop and Trout want reasoning strategies to achieve, in order for them to count as “excellent” or rational reasoning strategies?

Taking distance from Stich’s Epistemic Pragmatism, they state that a good outcome is nothing but a true belief. But they observe that people are not interested in all possible
truths, but, rather, in truth with respect to a class of significant problems (Bishop & Trout 2005: 93-95). Indeed, Strategic Reliabilism evaluates reasoning strategies not only in terms of their robust reliability and feasibility, but also in terms of the extent to which they lead to true beliefs about significant matters. Looking at the philosophical literature, you can find several ways to characterize significance. Among these characterizations, Bishop and Trout opt for a reason-based approach to significance. Their aim is to define a notion of significance that is thick enough to rule out the “significance” coming from preferences and interests to solve problems that are too idiosyncratic (e.g., someone might be interested in counting the precise number of grains of sand on a beach), but thin enough to acknowledge the obvious fact that there are interpersonal differences among people’s preferences and interests. Bishop and Trout (2005: 95) hold that “the significance of a problem for S is a function of the weight of the objective reasons S has for devoting resources to solving that problem”. Such a definition raises some questions: what is it for S to possess objective reasons for doing something? What are such reasons that prompt subjects to allocate resources to handle a certain problem? According to Bishop and Trout’s approach, what objective reasons S has does not amount to what reasons she thinks to have (subjective view) or what reasons someone else thinks she to have (interpersonal view): what objective reasons she has is a matter of what promotes her well-being (Bishop & Trout 2005: 99; Neta 2008). So, the significance of a problem is characterized in terms of the extent to which solving a certain problem promotes a reasoner’s well-being. The determining factor of what counts as a good outcome is its conduciveness to people’s well-being: hence, the good outcomes, which the Aristotelian Principle refers to, are significant truths relevant to our well-being.

Characterizing good outcomes in such a way, whether or not an allegedly reliable and feasible reasoning strategy is the excellent one compared to its competitors depends upon
the conduciveness of its outcomes, consisting of true beliefs, to well-being. Assume now that in a given circumstance a reasoning strategy leads to a certain outcome that is held to be conducive to well-being. Suppose that afterwards it is discovered that the outcome produced by the reasoning strategy is not in fact conducive to people’s well-being (e.g., what was formerly considered well-being turns out later on to be non well-being). Therefore, we can conclude that, after discovering that such an outcome (even if consisting of a true belief) is not conducive to people’s well-being, the reasoning strategy that led to it cannot be counted as an excellent one. As I understand their proposal, Bishop and Trout define what a good outcome is by referring to what people should or should not do with regard to the creation of conditions that promote their well-being. So, in my opinion, Bishop and Trout assume a normative conception of what counts as a good outcome that seems to be irreducible and lies outside the scope of naturalistic explanations. But if they really assume a normative conception of what counts as a good outcome, their Aristotelian Principle fails short of being, as they hold, a genuine empirical generalization.

3.5 Against Strategic Reliabilism

Bishop and Trout hold that Strategic Reliabilism provides a normative framework for assessing the different ways in which people reason on the basis of a series of pieces of empirical evidence coming from Ameliorative Psychology. But, as I will try to show, to the extent to which Strategic Reliabilism is a non-circular and empirically based proposal, its scope is limited and its alleged strong impact to the epistemological theorizing is therefore weaker than is claimed by Bishop and Trout.
3.4.1 The Aristotelian Principle

Insofar as Bishop and Trout (2005: 169) maintain that the Aristotelian Principle supports the inference that certain reasoning strategies are epistemically superior to their competitors, it seems to me that they assign to it a meta-normative role. But at the same time, and this is a questionable point, they claim that the Aristotelian Principle is grounded on an empirical basis, having introduced it as a “robust” empirical generalization. In other terms, that principle is held to be nothing but the outcome of an inductive generalization, and (it is alleged) of an obvious one, since “any psychologically healthy, reflective person who has chosen to spend their life doing epistemology must accept the Aristotelian Principle” (Bishop & Trout 2005: 20). So, on the one hand, the Aristotelian Principle allows “to take empirical results and infer from them with confidence that one way of reasoning is better than another” (Bishop & Trout 2005: 169), while, on the other hand, it has to be grounded on empirical observations about good or bad outcomes which competing reasoning strategies can lead to.

But how can an observer decide when the outcome of a particular reasoning strategy is good or bad? He would need some criteria so as to determine that. So, I would say, the definition of criteria of what is a good outcome is a necessary requirement for supporting the Aristotelian Principle empirically and those criteria have to be independent of the Aristotelian Principle itself, if appeal to it in assessing reasoning strategies is not to be vitiated by circularity. But the only factor determining what counts as a good outcome which may well be independent of the Aristotelian principle, that is, conduciveness to people’s well-being, is itself normative. So, we may conclude, the Aristotelian Principle’s reference to good outcomes introduces a normative aspect into it, which should hinder us from considering it as a mere empirical generalization on which inferences about which way of reasoning is better than another can be non-circularly grounded.
3.4.2 Well-being and good outcomes

In order to prevent bringing into the Aristotelian Principle a normative aspect, Bishop and Trout try to give “conduciveness to people’s well-being” a merely descriptive characterization reducing it to some basic descriptive conditions. But their attempt does not seem to be successful. Let me explain why. I begin with two questions: why should we attribute an intrinsic value only to truths which serve for our well-being? And which are the basic conditions of people’s well being? According to Bishop and Trout (2005: 99), “the basic conditions of human well-being involve health, deep social attachments, personal security, the pursuit of significant projects […]”. On the one hand, we can agree that any epistemic activity should attempt to increase (or at least not work against) the occurrence of conditions that promote people’s well-being. On the other hand, it seems to me that their characterization of well-being does not cover all the matters we might want to consider significant. The problem is that the truths that people would consider significant cannot be wholly laid upon those the possession of which may promote their well-being, as Bishop and Trout seem to hold. Many truths and achievements are related to the improvement of well-being in a rather indirect and rough way and, however, they still seem to be completely significant. For example, Bishop and Trout’s account seems to give to a reasoner no (objective) reason to care about foundational issues in a wide variety of academic disciplines, such as philosophy, natural science and mathematics. Clearly, only some, or perhaps none, of these questions may be related directly to matters of well-being. It is not clear however why, following Bishop and Trout’s theory, they should not be regarded as significant. So, the notion of significance, as specified by Bishop and Trout by referring to our well-being, appears to be very limited in scope. The characterization of people’s well-being put forward by them covers only some goals that people may have while engaging in reasoning strategies, so what will happen when people have other goals
or interests? It seems doubtful that significance can be directly and exclusively linked to human well-being, defined in the way Bishop and Trout do. What they should aim at is, if possible, a more complex and accurate analysis of the conditions of the well-being of people. Without this analysis, it seems doubtful that significance cannot be defined in the way Bishop and Trout do.

4. Cognitive and situational context

In this section, I discuss the implications that pragmatist approaches to rationality assessment, such as Epistemic Pragmatism and Strategic Reliabilism, have for what we have called the evaluative question (see Chapter 1, Section 1.1). In particular, I hold that the supporters of these approaches tend to neglect the distinction between the situational constraints and cognitive constraints under which reasoners operate and consequently fail to distinguish the respective roles of those constraints in rationality assessments. In my view, this failure is due to the lack of a clear distinction between general cognitive goals and specific goals depending on non-cognitive (situational) factors. That leads me to focus on the notion of situational context, which I characterize as complementary but distinct from the cognitive context assumed by pragmatists such as Stephen Stich. Finally, with regard to the evaluative question, I point out at the vacuity of any consequentialist approach not taking into serious account the role of situational constraints in human reasoning.

4.1 Pragmatist approaches to rationality assessment

The approaches of Epistemic Pragmatism and Strategic Reliabilism to reasoning performance assessment can be characterized as resource-relative forms of consequentialism based on pragmatist considerations (see Samuels et al. 2004: 168-172).
But why adopt such forms of consequentialism? What implications does their adoption have for normative and evaluative issues? At first sight, both these epistemological accounts provide several good suggestions that ought to be taken into consideration first when deciding which normative standards of rationality to adopt and then when applying them to a given reasoning performance. Their focus is both on what the reasoner’s goal is (the consequentialist component) and on what resources are available for achieving that goal (the resource-relative component). On the one side, they focus on the cognitive constraints under which reasoners operate, and of the resources available to them. On the other side, they characterize goals which people aim at in terms of one’s interests and needs, that is, what reasoners intrinsically value or attribute significance to. Two consequences can be drawn from that. On the normative side, whether a given reasoning performance is normatively appropriate or not depends on the reasoner’s cognitive situation, provided that the goal pursued by her is appropriate to her current interests, preferences, and cognitive resources. On the descriptive side, that means that reasoners usually set their goals taking into account the cognitive resources available to them. More generally, it is assumed that it is characteristic of any human reasoner to choose goals that are appropriate to the interests and resources one actually possesses: to make sense of a task requires fitting together the requirements of the task with what one wants and can do according to her cognitive situation. So, the value and appropriateness of a reasoner’s goal and their corresponding standards precede any rationality assessment. In my view, however, what it is needed for such an approach to rationality assessment is a closer examination of the manner in which goals, normative standards and cognitive resources interact with the non-cognitive context, or situational context as I shall here term it, wherein reasoning occurs.
Up to now, in this section, our discussion has been mainly concerned with what I shall call cognitive context, which can be identified with the reasoner’s preferences, interests, and cognitive assumptions. But, as is correctly admitted by both Stich and Bishop and Trout, assessments of reasoning performances must also be situation-dependent because the resources available to the reasoners, the reasoning strategies relevant for achieving what they want and the benefits and costs of using such reasoning strategies typically vary with what I have called the situational context (Stich 1990: 136-140; Bishop & Trout 2005: 71-78). However, when looking at their analyses, one can find that both Stich and Bishop and Trout pay little attention to and make little use of the situational context. Indeed, their interest for this kind of context is limited to its influence to the evaluation of costs and benefits in using a given reasoning strategy. Taking this aspect for granted, I hold that the situational context has a greater impact on reasoning performance. Not only the situational context influences the cognitive economy of a reasoner in using a given reasoning strategy, but also affects how she figures out what the reasoning problem she is approaching is about and consequently what kind of goals she has to aim at for solving this specific problem. Bishop and Trout are conscious of that. They regard the situational context as the set of cues which are relevant for solving the problem at hand. However, in Bishop and Trout’s examples, the problems which they refer to involve very definite kinds of social activities, such as diagnosing psychiatric disorders, predicting success or failure at school or in business and so on (Bishop & Trout 2005: 11-16), and consequently people already know what the problem is about and, in turn, what they have to do. The reasoners’ task is only to put the appropriate cues (which are given to them in numerical terms) into the appropriate formula to determine the correct solution. This kind of situations is not widespread in ordinary life. Reasoning problems do not usually begin with a clear set of information and cues; rather, most problems must be identified and reconstructed in context. Reasoning
performances are more complex matters than they might appear. To understand them, it is important to distinguish (i) the reasoner’s framing of the problem and (ii) the consequent adoption of a particular reasoning strategy to solve it. While Epistemic Pragmatism and Strategic Reliability focus on the costs and benefits of using one strategy or another, that is, their focus is on what happens at stage (ii), stage (i) is neglected by their analyses. So, they fail to consider the situational component of the evaluations of reasoning performances. As a consequence, what we should aim at is, if possible, a distinction between the roles of cognitive and situational constraints in human reasoning.

4.2 Cognitive and situation-dependent goals

In my view, the best way to highlight the roles of cognitive and situational contextual constraints in human reasoning is to take into consideration the fact that, when dealing with a reasoning problem, a reasoner behaves in a goal-directed way. According to Yama (2002), when performing a reasoning task, a reasoner aims at two different kinds of goals. On the one side, when a reasoning problem is given, the reasoner’s general cognitive goal is, according to her preferences and interests, to avoid too much effort and to spare her cognitive capacity in performing the task. That goal is characterized in terms of efficient allocation of cognitive resources to outcome-maximizing reasoning strategies. On the other side, given the existence of a reasoning problem in the context, the reasoner has to identify what the problem is about and sets her goal accordingly.

Not taking into serious consideration such a distinction may lead to misunderstandings about what rationality assessments are about. Indeed, every reasoning performance should be regarded as the result of the attempt of a reasoner to accommodate and achieve together the cognitive and the situation-dependent goals. It is a sort of trade-off between one’s cognitive economy and her need to handle reasoning problems. In order to achieve
cognitive goals, reasoners usually rely on the use of heuristics. *Availability* and *representativeness* heuristics can be considered as a central example of the cognitive tools employed by subjects and based on the limits of their cognitive system (for a characterization of such heuristics see Chapter 1, Section 3.4). As seen above (see Chapter 4, Section 3.2), such reasoning strategies are usually evaluated in terms of their reliability. However, as is said by Goldman (1986: 124), “reliability has nothing to do with goal-responsiveness”. While a reasoning strategy may be very reliable, it can be at the same time completely unresponsive to the subject’s current goals. So, current (situation-dependent) goals require normative considerations of a different kind than those related to reliability. In my view, these different kinds of normative considerations stem from the situational context wherein reasoning occurs. In fact, every reasoning problem is part of an embedding situation, which includes those parts of the context which are directly relevant to its solution. Although it seems to be silent, without giving explicit contributions to the reasoning task, the situational context can be regarded as constraining reasoning, that is, it can be regarded as a frame of reference which constrains the goals at which reasoning is directed.⁹

In Epistemic Pragmatism and Strategic Reliabilism, however, the distinction between the cognitive and situational contexts is blurred and consequently reasoning performances are evaluated mainly in terms of reasoner’s goals and cognitive resources. What both Epistemic Pragmatism and Strategic Reliabilism claim is the trivial point that cognitive context matters in reasoning, because it affects what questions one will aim at answering. However, putting the situational context aside leads Stich to conclude that there is not such thing as the appropriate representation of a reasoning problem relative to the situational context, but, rather, a multitude of representations which match the competing targets at which different

---

⁹ These considerations about the normative role of context will be developed in the next chapter (see Chapter 5, Section 1.4, 2.2).
reasoners may aim. As a consequence, not only one but many normatively appropriate answers to a reasoning problem are possible and that depends upon what are the reasoners’ goals and the cognitive resources available for achieving it. In my view, this is the core of the relativistic view put forth by pragmatists, such as Stephen Stich: reasoning is regarded as constrained by preferences, needs and interests of the reasoner without taking into serious consideration the situational context wherein reasoning takes place.

4.3 Beyond relativism: consequentialism and context-sensitivity

More generally, as to Stich’s relativistic claims, I think it is useful to distinguish between two different theses about rationality: (i) the claim that what it is rational to believe or do is relative to a particular situation and (ii) the claim that rationality is relative and so are rationality assessments too. Stich does not seem to make a clear distinction between these two claims. In my view, however, (i) can be regarded as a contextualist claim, and contextualism is not a particular case of relativism.\(^\text{10}\) While the latter holds that there are no universal standards of rationality, a context-based account of evaluation would claim that normative standards, although necessary, are not sufficient to yield correct rationality assessment as their applications are “irreducibly context dependent”. Consider the case of epistemology. While relativistic approaches to epistemology try to eliminate any reference to objectivity, contextualism, in its different forms (see, e.g., Cohen 1986; DeRose 1992), indexes assessment to various contextual and conversational factors without necessarily refusing objectivity altogether. For example, according to epistemological contextualism, different contexts define different epistemic standards, and thereby its supporters hold that standards may drastically differ from context to context, which is an objective fact after all.

\(^{10}\) This is why contextualist objections can be levelled at relativism (such as those I have put forth in Chapter 5, Section 2.5).
Consider now the question of the rationality assessment. According to relativist positions, any reasoning performance might be taken to be rational from a given cognitive situation. That is, by putting together a suitable set of cognitive assumptions about the problem that a subject assumes she is solving, surely her rationality can be defended. But this suggests the vacuity of any consequentialist approach which does not take seriously the situational component of the rationality assessment. Accordingly, in the next chapter, I will argue that a context-sensitive consequentialist approach is able to foreclose such a relativistic view of the rationality assessment. Reasoning is not only goal-directed, but also always situated. By including the reasoner’s cognitive state within a situational context, a contextualist approach would stress the importance of the reasoner’s ability to evaluate situations and select the relevant contextual information in order to successfully solve her reasoning problem. On this view, a reasoner ought to figure out what the problem at hand is about and what is relevant for solving it with reference to not only her cognitive situation, but also in accordance with the particular situation wherein reasoning occurs.

5. Concluding remarks

In this chapter, I have examined two consequentialist approaches to rationality assessment, both based on pragmatist considerations. While those accounts take reasoning strategy assessments to be constrained by the reasoner’s goal and the cognitive resources available for achieving it, they diverge in their characterizations of what counts as a good outcome: one the one hand, Epistemic Pragmatism focuses on the achievement of outcomes conducive to things which we (as individual reasoners) attribute an intrinsic value to and, on the other hand, Strategic Reliabilism focuses on the achievement of true beliefs about topics that the reasoner considers to be significant. Within a pragmatist framework, it is
held that there are no universal normative standards of rationality but, rather, any evaluation is dependent on the reasoner’s cognitive situation, with the *proviso* that the goal pursued by her be appropriate to her current interests, preferences, and cognitive assumptions. So, both Epistemic Pragmatism and Strategic Reliability tend to neglect the distinction between the situational and cognitive constraints under which reasoners operate and consequently fail to distinguish the respective roles of those constraints in rationality assessments. This failure is due to the fact that both these accounts miss a fundamental distinction between two different kinds of goals. Indeed, when dealing with a reasoning problem, although her cognitive system has always a general cognitive goal, the reasoner has to identify what the problem is about and set the goals of her reasoning activity accordingly. Epistemic Pragmatism and Strategic Reliability’s main interest is in the first kind of goal, as is shown by their focusing on benefits and costs in using one reasoning strategy or another. However, I have argued that assessing one’s reasoning performance only according to the reasoner’s own cognitive context may lead to relativistic claims about rationality, because any reasoning performance might be viewed as rational from at least one cognitive situation. That is why I have moved on to focus on the situational context. That leads me to consider the possibility of a context-sensitive consequentialist approach to rationality assessment. According to this approach, we treat the problem of assessing the rationality of a reasoning performance in terms of success in achieving one’s goals given the constraints imposed by the situational context. In the next chapter, I will argue for a situated view of rationality, according to which also the reasoner’s cognitive state is included in the situational context.
Chapter 5 – Towards a situated consequentialism

Introduction

In this chapter, I outline a way of approaching rationality assessment which I call the context-sensitive consequentialist approach and provide an example of how it can be applied to empirical data obtained from experimental studies based on reasoning tasks. As is well-known, in the last decades, different kinds of contextualist approaches have been employed in various philosophical fields, particularly epistemology, ethics, and the philosophy of language. Up to now, however, there has been no systematic attempt to apply such an approach to questions about human reasoning and rationality. Rather, researchers who have tried to explain away particular cases of alleged errors in reasoning by appealing to contextual and conversational factors (e.g. Hilton 1995; Politzer 1986; 2004), have failed to provide a general normative framework for their analyses. My attempt to develop a context-sensitive consequentialist approach to rationality assessment would like to be a way to fill this gap.

The chapter is divided into three main sections. In the first one, I characterize the general idea underlying the context-sensitive consequentialist approach to rationality assessment. I first examine the conversational approach to reasoning performance and the cognitive conception of context which its supporters assume. Then, I argue that such an approach has to be refined and integrated in order to provide a suitable normative framework for rationality assessment. In particular, starting from some considerations as regards the situatedness of speech acts, inspired by the work of J.L. Austin, I oppose an objective notion of context to the cognitive one assumed by conversational pragmatists, and then
explore the implications that the adoption of an objective notion of context would have for the development of an approach to rationality assessment. In the second section, I characterize the main assumptions and purposes of the context-sensitive consequentialist approach to rationality assessment and provide its building blocks. By focusing on the strong interdependence between the particular structure of a reasoning task and its understanding on the part of the subjects, I outline a two-step normative framework for situationally establishing the legitimacy of the subjects’ interpretations of the task and the normative appropriateness of their responses. In the last section, I consider how the context-sensitive consequentialist approach can be applied to empirical data obtained from a well-known experimental reasoning task, that is, the Wason selection task. After introducing an overview of the main issues related to the studies on this task, I examine whether in its standard version the most common answer is indeed as irrational as many scholars involved in the rationality debate have claimed. Applying the two-step normative framework, I maintain that although, relative to their ordinary goals, the subjects’ most common response can be considered as rational, their real error resides in their framing of the problem which does not match the information as explicitly presented in the experimental context.

1. Situational context and objectivity

In this section, I focus on the role of context in questions about human reasoning and rationality. First, I explain why reasoning cannot be detached from the context wherein it occurs. Second, I examine the conversational pragmatists’ approach to rationality assessment and criticize their cognitive conception of context. Third, I turn to John L. Austin’s analysis of the situatedness of speech acts and to the related idea of objective
Finally, I consider the implications that such a notion of context could have for rationality assessment.

1.1 Why does context matter in reasoning?

In the previous chapter, I have assumed that human reasoning invariably occurs within what I have called a situational context (see Chapter 4, Section 4). According to a strict cognitivist point of view, however, one could consider human reasoning as an independent and self-centred cognitive activity. For example, as is pointed out by Denis Hilton (1995: 248), “[m]ost psychologists conceive of judgment and reasoning as cognitive processes, which go on ‘in the head’ and involve only intrapsychic information processing”. Indeed, the processing by which a conclusion or decision is reached takes place entirely within people’s cognitive systems. Accordingly, the proponents of such a view would hold that cognitive processes underlying reasoning work in the absence of context, even if some contexts can adversely or positively affect some instances of reasoning. As Daniel Andler (1993: 291) observes, the idea that reasoning can occur in the absence of context calls for a clarification as to what “absence of context” means. On the one side, if you take this expression literally, it seems to be at least incoherent as any event takes place somewhere. On the other side, according to a more “technical” interpretation, “absence of context” can be understood as “something analogous to the notion of a physical process occurring in a vacuum” (Andler 1993: 291). According to Andler, many scholars would identify the vacuum where reasoning occurs with formality, namely absence of content. But this is a misleading equation. When faced with a reasoning task (even an elementary exercise coming from an introductory textbook of logic), people are attempting to solve it within a certain kind of context and formality itself is one aspect of the context. In Andler’s view, it
is clearly wrong to assume that reasoning may have a context-free mode, while sometimes is context-sensitive. According to him,

the confusion arises from thinking of ‘context-free’ as ‘occurring in normal, unmarked circumstances’; so context-sensitivity in the mild sense would amount to the existence of marked, non-default contexts leading to an output differing somewhat from the output obtained in the default context. (Andler 1993: 291)

Contextualization is at the core of human activity: whatever kind of activity people are engaged in, they are doing it within a context. Indeed, there is no such thing as solving a problem, deciding what to do and the like, outside of some context where the information, instructions and other similar cues are given. Accordingly, what is needed is a closer examination of the contribution of context to human activities such as inferring conclusions, making predictions, judging the likelihood of a particular event, making decisions, and testing hypotheses. As with other widely used notions that are commonly referred to in analyzing everyday activities, however, context is difficult to define and grasp in all its features and roles.

1.2 Context and conversation

In the study of human reasoning and rationality, the role of context has been widely discussed and analyzed by the supporters of the conversational or pragmatic approach to the analysis of reasoning performances (e.g., Hilton 1995; Politzer 1986; 2004; Politzer & Macchi 2000; 2005; Schwarz 1996; for a survey, see Lee 2006). Their pragmatic analyses are, above all, a reaction to the approaches to reasoning performance interpretation and evaluation employed in earlier psychological research on human reasoning. For a long time, researchers have maintained that subjects’ interpretations of the task have always to fit with
the experimenter’s representation of the task and so normative standards have to be appropriate to this representation, regardless of how they understand or interpret the task. It has been implicitly assumed that the task with which subjects are faced is a well-defined one, that is, it explicitly provides all information necessary to solve the task according to the representation the experimenter has assumed to be the right one. As is noted by Evans and Feeney (2004: 78), “any influence of prior knowledge or belief about the problem content or context [has been taken] to be normatively irrelevant to the definition of a correct answer”. Recently, however, these classical assumptions about the interpretation and evaluation of reasoning performances have been widely re-examined and criticized. Several researchers now claim that the evaluation of subjects’ performances on a reasoning task should be always relativized to their interpretation of the task and the conclusions they draw must be evaluated by considering both their goals and the background assumptions that they have selected as relevant to solving the reasoning problem (see, e.g., Evans & Feeney 2004; Girotto 2004). Such an approach invites a main worry: it could be the case that experimenters are too permissive with respect to the evaluation of subjects’ reasoning performance, that is, they may explain away any normatively inappropriate response by assuming that subjects have interpreted the reasoning problem in ways that are coherent with their responses. If they do so, experimenters deprive standards of rationality of all their normative force (see also Chapter 1, Section 3.5). To prevent that, the supporters of the conversational approach have proposed to explain reasoning performances by appealing to contextual and conversational factors. They consider reasoning as an activity which takes place in a context, be it either linguistic or interpersonal or both, and cannot be detached from it. On this view, the inseparability between reasoning and context is not only a practical concern, but also a theoretical stance.
Supporters of the conversational approach maintain that studying reasoning and rationality from the point of view of pragmatics allows us to discover new factors that are likely to determine subject’s reasoning performances (see Politzer & Macchi 2000; 2005). In their view, what received approaches have not sufficiently examined is the context which arises from the definition of a problem. Therefore, before any experiment on reasoning can be made, it is necessary to consider the possible ways in which subjects may understand and interpret the reasoning task they are faced with. As Politzer and Macchi (2005:120) observe, if after a reasoning problem has been analyzed from the point of view of pragmatics, it is discovered that subjects may have understood the reasoning problem in ways that differ from what the experimenter has assumed to be the case, and thereby subjects have approached a problem that has a different nature from that devised by the experimenter, that may have deep consequences on the assessment of subjects’ reasoning performances. In particular, according to Politzer and Macchi (2005: 120-121), experimental tasks for which a normatively correct response has been defined should be examined at two different levels:

One, carried out at a micro-structure level, consists of a linguistic analysis of the premises or of the problem statement in order to make sure that they convey the meaning intended by the experimenter. A typical outcome of such an analysis is the identification of different possible interpretations due to the generation of conversational implicatures […]. The other examination, at a macro-structure level, consists of identifying the representation of the task that participants are likely to build: a typical outcome of this examination is the identification of the kind of skill, knowledge, or ability that participants think they must exhibit in order to satisfy the experimenter’s request.
The second stage of analysis focuses on the relationship between experimenter and subject. Their relationship is taken to be asymmetrical because it is the subject who tries to understand what the experimenter’s intentions are. However, the intentions of the experimenter are not always completely transparent to the subject and thereby the latter may attribute intentions to the experimenter that may be very different from the experimenter’s expectations. If this occurs without being recognized by the experimenter, the subject’s interpretation of the task may affect the experimenter’s analysis of the results and her evaluation of the subject’s reasoning performance. Such a pragmatic method is situational in the sense that it focuses on situational (experimental) constraints in order to judge the normative appropriateness of subjects’ reasoning performances. The explanatory situation is constituted by the experimenter and her relationship with the subject. Opposing pessimism on human rationality, supporters of the conversational approach hold that, by appealing to contextual and conversational factors, subject responses may be said to be very often “conversationally rational”. As Hilton (1995: 264) points out, “many of the experimental results that have been attributed to faulty reasoning may be reinterpreted as being due to rational interpretations of experimenter-given information”. But what does it mean to be “conversationally rational”? What kind of context do conversationally inspired analyses of reasoning performances imply? At this point, I need to say something more about the nature of the context which is focused upon. Supporters of the conversational approach, as I understand their claims, define the context as the set of assumptions that the reasoner supposes herself to share with the experimenter. Indeed, as seen above, they regard the participant as trying to understand the experimenter’s intentions. These assumptions do not come from the situational context but rather are part of the subjects’ cognitive context.¹ If context amounts to the assumptions that the reasoner takes to be held

¹ As regards to pragmatic interpretation, a classical example of context regarded as “cognitive” comes from Relevance Theory (Sperber & Wilson 1995). In this theory, what counts as context is characterized in terms
in common by herself and the experimenter, then it is fairly internal and cognitive. On this view, a reasoning performance could be regarded as normatively inappropriate only in relation to the reasoner’s reconstruction of the intentions of the experimenter about the reasoning task she is faced with. Doing so, errors in reasoning may be always explained away by appealing to the reasoner’s cognitive assumptions and what we usually consider irrationality may become a conversationally rational way of reconstructing the goal set by the experimenter in the reasoning task. While such accounts are typically optimistic about human rationality, I hold that a contextualist account should not prevent criticism of subjects’ reasoning performances. Indeed, with regard to the same reasoning performance, subjects might be reasoning rationally in conversational terms, and yet be subject to criticism (for example, if the subject’s representation of the task context does not match the contextual information as explicitly presented in the experimental context). It seems to me that in order to provide an appropriate and complete evaluation of human reasoning it is necessary to consider the appropriateness of subjects’ task interpretations. But it is only with respect to something external to the reasoner and independent of her cognitive assumptions, that it makes sense to assess, or attempt to assess, her interpretations. So, if we want to have any normative standards for the evaluation of reasoning performances, then we must also have the means to evaluate the legitimacy of subjects’ task interpretations. To fulfil this need, the conversational account should be integrated into a more general normative framework.

1.3 Towards an objective notion of context

Let us consider the first step that has to be made in order to develop such a general normative framework. According to supporters of the pragmatic approach, as seen above,
the evaluation of the correctness of a response to a given reasoning task should be always relativized to the subjects’ interpretation of the experimenters’ intentions. As I understand their proposal, the reasoner’s cognitive context plays a fundamental role because the context is regarded as the set of possible assumptions that the reasoner supposes herself to share with the experimenter. But, if the context of evaluation corresponds to the cognitive situation of the reasoner, the appropriateness of a given reasoning performance would demand of the reasoner too little: it would demand of her that she behaves in accordance with her own interpretation of the situation but not that her understanding of the task displays a correct grasp of the situation. In my view, what is needed is something that is external to the reasoner’s cognitive context and independent of her cognitive situation in order to assess the legitimacy of her problem reconstruction, and I identify this “something” with the situational context by conceiving it as “objective”. The idea that the context of evaluation has to be regarded as objective has been held by few philosophers. Indeed, it is a controversial question whether we can delimit objective context completely. According to Carlo Penco, for example, if we are really interested in using an objective notion of context, we should integrate it into a cognitive one. He holds that

the objective context is, most of the times, the context we recognize as objective.

We know both that there is some objective reality and that we might get it wrong. To describe an objective context as such, independent of a cognitive one, is therefore a risky enterprise. Any attempt to define it in an absolute way is misleading, because it takes a description – given always inside some theory or cognitive context – as an objective unrevisable description. Objectivity is always a result of our interaction, not a datum […]. (Penco 1999: 280)

I admit that identifying the context of a conversational event in an absolutely objective way is a risky (or maybe an impossible) enterprise. However, objective context may be
characterized without requiring such an absolute point of view. Most prominently, as pointed out by Marina Sbisà, the idea that the context of evaluation should be regarded as objective plays a fundamental role in the evaluation of speech acts, as originally characterized by John L. Austin (Austin 1975; Sbisà 2002). Without entering into the details of speech act theory, I focus on the role that Austin attributes to the situational context in the evaluation of an assertion as true or false. According to Austin (1975: 143), although an assertion such as “France is hexagonal” is usually considered to be perfectly determinate, it cannot be said to be true or false until the interlocutors’ goals are specified, which usually happens tacitly. As a consequence, in order to qualify an assertion as true or false, the context in which it has been made, as well as the interlocutors’ goals, has to be taken into account. So, the assertion “France is hexagonal” may be judged true if made in a certain context with a certain goal (i.e., a general considering the sides from which his army could invade France), but may be judged false if made in another context with a different goal (i.e., a geographer describing the borders of France in detail) (Austin 1975: 142; Sbisà 2002: 426). Within this framework, it is assumed that the goals of the interlocutors determine the aspects of a situation against which the truth/falsity of a speech act concerning that situational context is to be evaluated. So, the fact that the truth or falsity of a sentence may vary from context to context shows that the situatedness of the assertion (like that of any other speech act) is strictly linked with the delimitation of its context. As Marina Sbisà (2002: 427) points out,

2 It is noteworthy that, according to Austin (1975: 141-144), truth/falsity is one among many other criteria for evaluation of a speech act. In contrast with mainstream philosophy of language, he also holds that truth/falsity is often not the best choice for evaluating an assertion. However, Austin’s claims have nothing in common with Stich’s attack against truth as our main cognitive goal (for Stich’s argument see Chapter 4, Section 2.2). Austin does not hold, as Stich does, that there is not only our idiosyncratic concept of truth, but many others, and which is the best one to rely on depends on its effectiveness to achieve things we intrinsically value. Austin’s approach is neither pragmatist nor relativist (Austin 1975: 144).
if a speech act is produced and understood in a context and is therefore a situated event, it seems reasonable to think that it should be evaluated with respect to that context. Consequently, in order to yield a definite evaluation of a speech act (in terms of felicity/infelicity, appropriateness/inappropriateness, truth/falsity), context must itself be delimited [...].

Three interesting lessons can be drawn from Austin’s view of speech acts: (i) every speech act, being produced and understood in a context, is a situated event; (ii) its context of evaluation corresponds to neither the participants’ cognitive contexts nor the contextual assumptions that they suppose to share with one another; rather, it depends on the context in which the interlocutors are situated; (iii) the delimitation of the context of evaluation is determined by the interlocutors’ goals. In my view, the kind of normative framework Austin proposes for the evaluation of an assertion as true or false has interesting implications for the development of a context-sensitive approach to rationality assessment.

1.4 Rationality assessment and objective context

Consider now how Austin’s lesson is relevant for the question of rationality assessment. First of all, I am conscious that speech acts and reasoning performances are two different kinds of human behavior, albeit sometimes overlapping in practice. But they share two fundamental characteristics, that is, they are both situated and goal-oriented. That has led me to wonder whether instances of reasoning, as well as speech acts, should be appropriately evaluated according to the context wherein they occur and the goals in the light of which they are performed. Consider the roles that goals and context play in rationality assessment. On the goal side, different purposes are served by different kinds of reasoning. In consideration of that, reasoning performances should not be assessed in abstraction from their different purposes: when assessing a reasoning performance, we
should first ask what purpose reasoner’s answers are aimed at and then consider whether they are right or wrong, correct or incorrect with respect to it. Indeed, without any previous reference to the goal pursued by the reasoner, assessing her reasoning performance seems to be pointless. On the context side, when an individual is about to engage in some activity, such as performing a task or solving a problem, there is an objective situation that determines the type and quality of information actually available. This situation should not be regarded as part of the task or problem at hand, but rather as what generates the task or problem on the subject’s perspective. Without situational context, no reasoning task or problem can occur. However, a main worry with the situational context is that there are no clear criteria for its delimitation while it has to be limited. In order to account for the delimitation of context, the most natural candidate may be the fact that not all the situationally available information is actually relevant to the goal the reasoner is trying to achieve (a suggestion in this direction comes from Gauker 1998; 2003: 55-58). In particular, the reasoner’s goal can be regarded as that which delimits the situational context, distinguishing the situationally available information which is relevant from that which is not relevant to reach it. Once the situational context is set by the reasoner’s goal, the problem she has to tackle is in principle defined and its solution is determined (about which, however, the reasoner may be wrong in many ways). Thus, the result of the interaction between the situational context and the reasoner’s goal gives rise to a frame of reference which constrains the ways in which the reasoner’s goal is to be attained. In other words, such a delimitation of the situational context leads to a normative frame, regardless of how the subject conceives of it.³ One may object that if the goal the subject aims at contributes to the delimitation of the relevant normative frame, such a frame is in part subjective. In my view, however, the fact that the goal of the subject plays a fundamental

³ For a different characterization of the normativity of context see Andler 2000.
role in the delimitation of context does not keep the resulting framing of the problem from being subject to normative evaluation. Starting from the same objective situation and goal, one thing is how the reasoner conceives of the frame within which he is trying to achieve the goal and another thing is the normative frame which dictates the ways in which the reasoner’s goal has to be attained, and about which the reasoner may be wrong in many ways.

What I have presented here is only a rough account of how goal and context might come into the rationality assessment. I have assumed both that reasoners have goals and that there is a distinction between those instances of reasoning that fit with a context and those that do not fit. Of course, greater effort has to be put into providing a more detailed account of how such a contextualist approach can be applied to specific situations, if we want to make a serious attempt at making the consequentialist picture of human rationality fully situated.

2. A context-sensitive consequentialist approach to rationality assessment

In this section, I characterize the building blocks of the context-sensitive consequentialist approach to rationality assessment sketched out in the last part of the previous section and explain in particular how the resulting approach can be applied to data on human reasoning obtained in experimental settings. First, by focusing on the particular structure of a reasoning task and its understanding on the part of the subjects, I provide some preliminary considerations as to how people approach reasoning problems and then point out the consequences that these considerations have for an approach to rationality assessment. Second, I propose a two-step normative framework for situationally establishing the legitimacy of subjects’ task interpretations and the normative appropriateness of their responses. Finally, I explain why such a normative framework can both account for the
situatedness of human reasoning and provide a satisfactory normative background against which to assess it.

2.1 Goals, relevant information and reasoning strategies

To begin with, it is clear that the subject’s answer to a reasoning problem is not self-explanatory, namely her answer itself does not give any explicit advice as to how she has arrived at it. So, if we want to assess how people reason, we have to move beyond the output of their strategies in reasoning tasks. In simple terms, it is impossible to assess a piece of reasoning as rational or irrational without first understanding it. Addressing this issue requires that we dig below the output surface. The outputs of people’s reasoning strategies are the tip of an iceberg. It is not my aim here to discover and explain what is below the waterline, that is, develop a model of cognitive performance in reasoning tasks. However, in attempting to develop an approach to rationality assessment, the first task is to delimit and define better what counts as a reasoning performance. That is, we have to make explicit what is subject to assessment. Only when the object of evaluation has been identified, we can start setting the appropriate normative standards for it.

When analyzing empirical data on human reasoning, most cognitive psychologists have assumed that every reasoning task is associated with a single normative model which has been considered to be not only the benchmark of normatively appropriate performance in the reasoning task, but also the interpretive grid of that performance. In other words, normative models of rationality have been assumed in order to both articulate normative standards and describe reasoning performances. However, one task is to explain why subjects’ answers are made and another one is to assess their rationality. Accordingly, these two tasks have to be neatly distinguished. Otherwise, misunderstanding may arise about what subjects are really doing in a given task. To understand a reasoning performance, it is
important to single out (i) the reasoner’s framing of the reasoning problem (i.e., the
reasoner’s understanding of what the problem is about) and (ii) the particular reasoning
strategy which she adopts to solve it. Stage (i) has been usually regarded as not belonging
to reasoning performance. As is showed by proponents of the conversational approach,
however, it is an undeniable fact that the reasoner’s framing of the problem determines the
decision about which reasoning strategy to adopt. As a result, it could be argued that the
subjects’ understanding of the problem determines to a certain extent the outcome of their
reasoning. It is a matter of fact that problem structure and subjects’ understanding of it have
a bearing on the question of rationality assessment.

Let us consider more deeply the structure of experimental reasoning tasks, in which all
relevant information is available in the experimental context, and the situation in which the
problem is raised does not change over time. If we look at such tasks, we find that in the
great majority of cases, problem presentations comprise a set of information (the premises,
the context surrounding the premises, the instructions, the examples...) and a question,
albeit one that is not fully explicitly formulated. When reasoning experiments are
presented, subjects primarily deal with a task of discovery, that is, the task of discovering
what the problem is about and what goal they are required to achieve. In other words,
before all, subjects have to orient themselves. Questions that might be asked on their part
are: what kind of task am I faced with? And then, what should I do with this particular
case? To fulfil these questions, subjects fix a goal according to their interpretations of the
question posed by the experiment and work towards it. Once the goal is determined, they
usually frame the problem in a way that makes sense of part of the situationally available
information: so, (i) subjects may ignore a piece of information or consider it to mean
something else and (ii) they may also bring other information (retrieved from memory) into the problem.4

With regard to (i), various cases may occur. Firstly, as is pointed out by supporters of the conversational approach, subjects may interpret information obtained from the problem presentation and the experimental setting in ways that the experimenters may not have considered and that conflict with their interpretation of the problem. As a result, information, which is taken to be irrelevant according to the experimenters, may be taken to be relevant by the experimental subjects. Such allegedly relevant information may change the nature of the reasoning problem and thereby what the normatively appropriate response to the problem is (see, e.g., Sperber et al. 1995: 44). Secondly, in experimental contexts as well as in ordinary life, subjects do not necessarily take into consideration all the information available in the context. They may consider some pieces of information to be more relevant than others and, given the limitation of their cognitive capacities, may tend to use their time and effort for information from which they expect they could benefit (van der Henst 2006).

As to (ii), subjects bring into the current situation their experience with similar problems, such as their knowledge related to the question posed by the problem. Sometimes that will lead to the recombination of contextual information in a way that changes the framing of the problem from how it was intended by the experimenters. Finally, all of the information that has been identified as relevant must be represented in a format that can fit with a reasoning strategy and this strategy will be applied to the represented information in order to reach the reasoner’s goal. Even when a problem has been framed, there always remains the question of which reasoning strategy best applies in such a case. In order to reach the reasoner’s goal, different reasoning strategies can be selected: in the case of

4 Keith Stanovich (2003: 292) has argued that “the tendency to automatically bring prior knowledge to bear when solving problems” is so widespread that it cannot easily be blocked. Consequently, he has characterized this tendency as “the fundamental computational bias of human cognition” (see also Stanovich 1999).
hypothesis testing, for instance, one can adopt either verificationist or falsificationist reasoning strategies depending on the information available in the context.

Distinguishing between the subject’s understanding of the reasoning problem and her choice of a particular reasoning strategy to solve it helps us understand how the context of the problem is mentally recognised and examined and why people could find the problem difficult. I am not addressing here the psychological question of how the information provided by the problem presentation is mentally processed and represented, or how cognitive processes underlying reasoning can be characterized. I would like to remain neutral about that. My primary aim here has been to show that, because of its complexity, assessing a reasoning performance requires a more complex and refined normative background than those usually assumed. The comprehension of the task and the search for relevant information should be regarded as part of every reasoning performance (see, e.g., Sperber et al. 1995: 44). Focusing on such aspects should help us define what task subjects are attempting to perform and consequently what kind of a normative standard should be adopted and applied to them in assessing their performances.

2.2 A two-step normative framework for rationality assessment

The last section has been aimed at giving an overview of the factors which influence and determine any reasoning performance. In accordance with the picture given, I would like to provide a two-step normative framework that is applicable to both subjects’ task interpretations and responses. In the vast majority of reasoning studies, it has been taken for granted that the selection of a normative standard for the assessment of a subjects’ reasoning performance is completely distinguished from the question of the subjects’ interpretations of the reasoning task. Thus, subjects’ response accuracy has been regarded

---

5 As is said by Evans and Over (1996: 9), “people are usually aware of a goal they have and may have some awareness of some of the steps required to attain it. But they are usually unaware and unable to describe the processes that do much to help them to achieve their goals”.
as the only object of evaluation in rationality assessment. On this view, rationality and reasoning have nothing to do with subjects’ understanding of the task and their search for relevant information, which have been regarded as mere cognitive variables. As seen in the previous chapters, reasoning performances have been assessed in terms of one’s ability either to comply with a set of normative principles (in the case of deontological approaches) or to reach a specific range of goals efficiently (in the case of consequentialist approaches). However, as suggested in the previous section, no normative claim about specific pieces of reasoning should be made without previously examining how subjects have framed the reasoning problem and what background assumptions they have made. Subjects may have different possible “understandings” of the problem context and the one they select influences how they approach the reasoning problem and what answers are the “correct” ones to give. As well as subjects’ responses to tasks, it seems to me that subjects’ task interpretations require a normative background against which to be assessed as to appropriateness or legitimacy. Otherwise, not only one but many normatively appropriate answers to a reasoning problem are possible and which one should be selected depends upon what the reasoners’ understanding of the problem is. Any response should be regarded as normatively appropriate directly following from the subjects’ representation of the problem and its context. By assuming such a view, however, rationality assessments become so flexible that whatever subjects do, they may always be regarded as rational.

What is needed is half-way between classical approaches to rationality assessment and relativistic ones. In my view, such a half-way approach involves a two-step normative framework for situationally establishing the legitimacy of subjects’ task interpretations and the normative appropriateness of their responses (for a similar analysis see Todorov 1997). This approach specifies two kinds of constraints that operate within the experimental context. On the one side, the subjects’ understanding of the reasoning problem is always
constrained by the information explicitly presented in the experimental situation. On the other side, as is held by consequentialists, the evaluation of success in reasoning cannot be separated from the evaluation of success in achieving goals. Let me explain these two points.

On the “understanding” side, before all, we need examine the presentation of the problem, which strongly influences the subjects’ understanding of it. When an individual is about to engage in some activity, such as performing a task, there is an objective situation that determines the type and quality of information actually available. For an experimenter, the first step is to engage in a descriptive exploration of the range of interpretations that the reasoning problem admits and their context-dependence. Otherwise, there is a risk that experimenters misconceive the subjects’ reasoning performances. Consider two examples. With regard to the Wason selection task, the semantics of conditionals suggests that there is more than one possible interpretation for statements such as those which Wason used in the selection task, envisaging the conditional as representing different conditional relations (Stenning & van Lambalgen 2001; 2004). Over the years, several of these interpretations have been adopted in order to explain subjects’ selections. The question is how to decide which interpretations may legitimately apply to the reasoning problem as proposed by the experimenter. In the same vein, depending on the context, words such as “probability” and “likely” may be interpreted in very different ways which do not strictly match their mathematical meaning. In the case of the Linda Problem, Jonathan Adler (1991: 261) has demonstrated that different meanings of probability may “all legitimately apply to the problem as posed, and the evidence does not decisively show one of these to be uniquely applicable”. More generally, as shown by proponents of the conversational approach, subjects do not usually have an explicit and complete grasp of the problem as intended by

---

6 Gerd Gigerenzer (1996: 593) has observed the word “probable” can mean, for example, “plausible”, “having the appearance of truth” and “that which may in view of present evidence be reasonably expected to happen”.

182
the experimenter, but most of them certainly have an implicit grasp of what the problem is about and that allows them to construct, with suitable contextual support, a wide range of interpretations of it. If they do not find a set of clear cues or instructions, the subjects try to frame the problem in the light of their previous ordinary and common experience. Indeed, insofar as all what they need is to understand what kind of task they are faced with and to figure out what they have to do in that particular situation, the subjects tend always to interpret the task giving to it practical significance (usually strictly connected with their ordinary activities) (see, e.g., Stanovich 1999: 190-207). Regardless of how experimenters interpret a reasoning problem, if the subjects’ framing of the problem matches the contextual information available in the experimental situation, it should be regarded as a legitimate interpretation of that problem. Thus, the same problem may be framed in different ways on the condition that the resulting representation matches the contextual information as explicitly presented. For example, as we will see with respect to the selection task, when dealing with a reasoning problem, subjects usually make several background assumptions about how things actually are in the light of their previous experience so that the normatively appropriate response they draw can be due to their reliance on these assumptions rather than to the use of a particular reasoning strategy. Generalizing from these considerations, I hold that the first set of constraints on rationality assessment comes from the experimental setting which delimits the range of frames which subjects may legitimately apply to a problem.

I turn now to the goal-relativity of the normative framework. Depending on the goal-structure of any framings, in order to assess reasoning performances we should first ask what purpose reasoner’s answers are aimed at and then consider whether they are right or wrong, correct or incorrect with respect to it. What does the subject aim at, and what means will she use to achieve it effectively? Such questions call for a consequentialist approach.
People show their rationality also in achieving simple and basic goals in ordinary activities. Contrary to the claims of evolutionary psychologists and pragmatists, I am not referring to the most important goals of one’s life, or to reproductive success. Indeed, our long-term goals do not always meet our short-term decisions, which are usually made within a specific situational context. Our general goal of good health may be in conflict with our preferences for fat foods and cigarettes. Goal-directed activities are as different and varied as the whole range of human activities. In an experimental setting, subjects may engage in an activity such as inferring conclusions, making predictions, judging the likelihood of a particular event, making decisions, testing hypotheses and so on. In doing so, subjects are not necessarily attempting to satisfy the experimenters’ expectations, but rather to solve the problem as they understand it. Thus, the correctness of a reasoning performance is determined in terms of its efficiency and effectiveness in attaining one’s goals. Indeed, it is an obvious fact that in any goal-oriented activity some ways of achieving goals are better than others.

To sum up, the two-step normative framework of the context-sensitive consequentialist approach states that the information explicitly presented in the experimental context places constraints on the goals and frames that it will be appropriate for the reasoner to activate when dealing with the task, which in turn determine the type of reasoning strategy that will be effective and efficient.

2.3 Situatedness and normativity

Let us consider more attentively the role to be attributed to the subjects’ understanding of the reasoning problem in rationality assessment, and how strictly it is related to how experimenters conceive of the situational context.
The subjects’ understanding of a reasoning problem has always been the crux of the researchers involved in the rationality debate. Two different tendencies can be found. On the one hand, those who trace subjects’ reasoning performances back to a predetermined normative standard assume that their subjects have understood the reasoning problem in the same way as they do and with the same background assumptions. However, as seen above, this approach is flawed as the same reasoning problem may be legitimately interpreted in different ways. On the other hand, other researchers hold that there is always an appropriate way to interpret the answers of the subjects as showing that they have understood the reasoning problem differently from how it was understood by the experimenter, so that their answers can be regarded as normatively appropriate to that which they respond to. In this second case, the priority is given to observed reasoning performance, not to the normative model: when assessing reasoning performances, the adoption of a normative principle has to be situationally justified in the light of what people actually do. However, the flexibility of what may be viewed as normatively appropriate may appear to raise a fundamental problem for the entire approach. As I have argued, it is only with respect to something external to the reasoner and independent of her cognitive assumptions, that it makes sense to assess, or attempt to assess, her reasoning performance. Accordingly, I have proposed an approach to rationality assessment according to which if reasoning is a situated activity, being performed in a context, it seems reasonable to think that it should be evaluated with respect to that context. Such an approach allows us to see reasoning performances, which have typically been assessed in strictly cognitive terms, as having a complex relation with the context wherein they take place. As said above, context should not be regarded as a set of reasoner’s cognitive assumptions but, rather, as something objective, determined by how the actual situation is and by the goals of the ongoing activities. In contrast with the two approaches mentioned above, before one can
legitimately assess an instance of reasoning, one must know the frame within which the reasoning is being done. Within this theoretical framework, to be rational means to do the best one can do in the actual circumstances, and the actual circumstances are where the reasoner is situated. Given her goal, the reasoner ought to make the best use of all the relevant information at her disposal.

With regard to this point, consider the Linda Problem again (see Chapter 1, Section 3.1.2). There has been considerable debate about Tversky and Kahneman’s interpretation of the empirical data they have obtained from this experimental reasoning task (see for a summary, e.g., Stanovich 1999: 121-124). In particular, many of their critics have argued that there are alternative interpretations of the problem that are, given the problem context, more appropriate than that which Tversky and Kahneman regard as the appropriate one. But what kind of goal subjects think is required of them to achieve in the Linda Problem? As Politzer and Macchi (2000: 87) observe, “since they are requested to produce a judgment of probability on the basis of the description of a character, in all likelihood participants wish to show that they possess the skills to find what maximizes psychological and behavioural coherence by identifying the kind of activity which provides greater relevance to the description of the character”. If this is really what subjects aim at in tasks based on the Linda Problem, we may call into question the idea that they are reasoning in a normatively inappropriate way: given the way in which subjects understand the reasoning problem, they have not contravened any normative principle of rationality. If we assume that subjects, after having read the problem presentation, make the supposition that the total amount of information given about Linda suggests that the experimenter knows a lot of things about her, it becomes reasonable to understand the statement “Linda is a bank teller” as implicitly conveying that she is not active in the feminist movement. Obviously, if subjects understand the statement “Linda is a bank teller” in such a way, then the fact that
they rate “Linda is a bank teller and is active in the feminist movement” as more likely than “Linda is a bank teller” cannot count as an error (Hilton 1995: 260; see also Dulany & Hilton 1991). Given all the information available in the experimental context, the subjects’ interpretation of the task does not seem to be inappropriate, so that subjects can be considered as making the best use of all the relevant information at their disposal.

3. Applying the context-sensitive consequentialist approach: the case of the Wason Selection Task

In this section, I consider how the context-sensitive consequentialist approach to rationality assessment can be applied to examine data on human reasoning obtained from a well-known experimental task, that is, the Wason selection task. In my view, the ways in which this task has been studied and discussed have strongly influenced discussions about human rationality. In particular, I will consider here the central question of whether in its standard version the most common answer should be considered as evidence of human irrationality as many psychologists have claimed.

The section is divided into four parts. In the first part, I reconsider the standard version of the selection task by focusing on Wason’s use of the hypothetico-deductive model as normative standard against which to assess subjects’ selections. In the second part, I provide an overview of the researches based on the selection task. In the third part, I point out to the differences between descriptive and deontic versions of the task and explain why in debating on human rationality descriptive versions give rise to greater problems than deontic versions do. I then turn to two attempts to explain subjects’ performances on selection tasks with indicative conditionals and discuss them, proceeding to a critical assessment of them. Applying the context-sensitive consequentialist approach, I maintain
that although, relative to their ordinary goals, the subjects’ most common response can be considered as rational, their real error resides in their framing of the problem which does not match the information as explicitly presented in the experimental context. Indeed, as we will see, most of the subjects tend to evaluate the indicative conditional with reference to a wider domain of evidence than the narrow one determined by the experimental context, as they do in ordinary life.

3.1 Standard version of the selection task and hypothesis testing

In the first chapter, I pointed out that Peter Wason designed the selection task as one of hypothesis-testing (see Chapter 1, Section 3.1.1). In Wason’s view, this kind of task calls for deductive reasoning based on a strictly logical interpretation of conditionals. As Oaksford and Chater (2002: 197) observe, “the assumption that the selection task is deductive in character arises from the fact that psychologists of reasoning have tacitly accepted Popper’s hypothetico-deductive philosophy of science”. Roughly speaking, Popper held that in empirical research the relevant observations are those that falsify, not those that confirm, scientific hypotheses. And a scientific hypothesis is falsified when predictions that can be logically drawn from it do not accord with empirical observations. As a consequence, Popper maintained that scientists should aim at designing experiments which can provide evidence falsifying the hypothesis under examination. Similarly, if experimenters apply the hypothetico-deductive method to the selection task, the only normatively appropriate selection consists of checking for the cards that may provide a falsifying instance of the conditional statement. As is said by Oaksford and Chater (2002: 197), “when viewed in these terms, the selection task has a deductive component, in that the subject must deduce logically which cards would be incompatible with the conditional statement”. Since the subjects’ most common choice (the combination of $p$ and $q$ cards)
only confirm the conditional statement, Wason (1968) concluded that most subjects display what he called a verification (or confirmation) bias, that is, they looked for instances confirming the conditional rule, and neglected instances falsifying it. So, it is Wason’s commitment to the Popperian hypothetico-deductive model that led him to refuse confirmation as an appropriate strategy in the selection task.

3.2 A brief history of the studies on the selection task

In this section, I present just a few highlights of the history of the studies on the selection task, focusing on some of its modified versions in which most subjects make the correct selection ($p$ and not-$q$ cards). The results of these researches show that subjects’ selections vary in terms of how the problem presentation is formulated. In particular, when the selection task is framed in certain ways (about which we will speak in more details in this section), we can expect that most subjects (usually over 70% of them) will give the correct response.

Experiments based on the standard version of the selection task were replicated more than once and the data obtained were almost the same (see, e.g., Wason & Johnson-Laird 1972), confirming thereby the robustness of Wason’s findings. However, Wason and Shapiro (1971) found that using thematic content in the task helps subjects to make the correct selection. In an experimental study subjects were asked to verify the conditional statement “If I go to Manchester [$p$], then I travel by car [$q$]” by examining four cards with the city destination on the one side and the transport used on the other side. The visible faces of the cards showed respectively: “Manchester” ($p$), “Leeds” (not-$p$), “Car” ($q$) and “Train” (not-$q$). The experimental results showed that this version of the task elicits a greater number of correct responses, that is, the $p$ (Manchester) and not-$q$ (train) cards (Wason & Shapiro 1971: 68). In order to explain this result, Wason and Shapiro propose a
hypothesis, which Griggs and Cox (1982) have subsequently dubbed “thematic facilitation effect”, according to which the use of realistic material facilitates subjects to consider different combinations of cards in order to identify the normatively appropriate solution.

Consider another modified version of the selection task. In a task devised by Philip Johnson-Laird, Paolo Legrenzi and Maria Sonino Legrenzi (1972), subjects were asked to imagine they were Post Office workers who had to check envelopes for violations of the conditional rule “If a letter is sealed [p], then it has a 5d stamp on it [q]”. Four envelopes (instead of the usual cards) were presented in front of the subjects. Subjects could see the back side of two envelopes and the front side of the other two. As to the first two, one envelope was sealed (p) and the other was not (not-p). The other two had respectively a 5d stamp (q) and a 4d stamp (not-q) on their visible faces (both envelopes had an address printed on them). Approximately 90% of the subjects made the correct choice: they selected the sealed envelop (p) and that with a 4d stamp (not-q) (Johnson-Laird et al. 1972). This finding was confirmed by subsequent studies using both Italian and British stamps with their respective units of currency (in both cases, the subjects were English). Johnson-Laird and his colleagues interpreted these results as confirming the “thematic facilitation effect” hypothesis. But subsequent studies have shown that using realistic materials does not always elicit the correct response in the selection task (see, e.g., Manktelow & Evans 1979; Pollard 1982).

In a replication of the experiment based on the postal task, Griggs and Cox (1982: 411-414) discovered that most American subjects tended to select the sealed envelope (p) and the one with the stamp that corresponds to that cited in the conditional rule (q) (experimenters replaced the British stamps with American ones). Griggs and Cox held that the significant differences between choices made by British and American subjects was due to the fact that, while the former had had direct experience with that type of postal rule
(indeed, a similar postal rule existed in the British postal regulation before the seventies), American subjects had never experienced that rule before because there was no postal regulation in the United States concerning amount of postage and the sealing of envelopes (Griggs and Cox 1982: 417). This explanation was confirmed by an experimental study made by Evelyn Goldman, who presented the same task to a group of British subjects who had never met the postal rule cited before because of their ages (indeed, this type of postal rule was eliminated in the seventies in Great Britain). The result of Goldman’s experimental studies showed that only few subjects gave the correct response (reported by Griggs and Cox 1982: 418n). In order to account for such results, Griggs and Cox (1982: 417) proposed the “memory-cueing” hypothesis, according to which in certain versions of the selection task people give the correct response because they can retrieve from their memory relevant counter-examples to the rules to be tested.

Subsequent studies showed that the improvement in subjects’ performance has to do with the nature and structure of the task, not with its thematic content or with its degree of familiarity (see, e.g., Manktelow & Over 1991; 1995). In particular, some experimental studies have shown that when asked to reason about certain kind of rules or regulations, most subjects give the normatively appropriate response. A well-known example of this kind of tasks is the “drinking age problem” (Griggs & Cox 1982). In this task, subjects were asked to pretend to be police officers being in a bar and checking whether the following conditional rule is being obeyed:

If a person is drinking beer \([p]\), then the person must be over 19 years of age \([q]\).

In this experiment the cards represented drinkers, showing the drink on one side and their ages on the other side. The visible sides of the cards were: “Drinking a beer” \([p]\), “Drinking coke” \([\neg p]\), “16 years of age” \([q]\) and “22 years of age” \([\neg q]\). The correct choice is to turn over the cards whose visible sides are “drinking beer” \([p]\) and “16 years of age” \([\neg q]\).
Griggs and Cox (1982: 414-417) found that around 75% of the subjects made the correct selection. While they thought that this result confirmed their “memory-cueing” hypothesis, other studies have demonstrated that this and other similar versions of the task elicit good responses because of their logical structure. In particular, subjects drastically improve their performances when the task is framed in such a way that what they are asked to check is concerned with permissions, prohibitions and obligations (see, e.g., Manktelow & Over 1991). Starting from these empirical results, many researchers have argued that people are good at reasoning with deontic conditionals because they possess domain-specific cognitive mechanisms which are specialized to handle with permissions and obligations (see, e.g., Cheng & Holyoak 1985; Cummins 1996). As seen in the third chapter, Leda Cosmides and John Tooby have proposed the so-called Cheater-Detection Hypothesis in the context of their evolutionary perspective on human mind (Cosmides 1989; Cosmides & Tooby 1992; see Chapter 3, Section 2). According to them, the empirical data obtained from experiments based on the selection task show that subjects have a particular mental module whose domain of application is restricted to conditional rules that involve the detection of cheaters. In their view, in the “drinking age problem” an underage drinking beer can be taken to be a cheater. So, the structure of the task activates the mental module for cheater detection, leading to the correct selection of the cards. It is noteworthy that, according to Cosmides and Tooby, the module for detecting cheaters can be applied to all the versions of the selection task which involve social regulations, but not to the standard one. Accordingly, they hold that correct performances in this kind of tasks should not be attributed to a strictly logical interpretation of the task.

3.3 Focusing on the indicative version: comprehension and probability

It is widely agreed upon that in the selection task participants improve their performances when faced with conditional statements expressing obligations or permissions. On the
contrary, most subjects fail to check for the potential counter-examples “p and not-q” when evaluating indicative conditionals. Why do people do better in the deontic version of the selection task than in its indicative version?

Setting aside evolutionary psychologists’ hypothesis, I would like to focus on a different sort of explanation that can account for why people reason better when faced with deontic versions of the selection task (see, e.g., Botterill & Carruthers 1999: 121-122). Indicative and deontic versions of the selection task have different logical structures, they present two different types of problems which include different kinds of conditionals and instructions. In the standard indicative versions of the task, subjects are asked to choose the cards which show whether the conditional statement is true or not. In the deontic version of the task, instead, subjects are asked to select the cards which show whether the conditional rule is being violated, as opposed to checking whether it is true or false. In the first case, subjects are explicitly told that the conditional statement may be true or false and so they are not allowed to assume that the conditional statement is true. In the second case, the truth or falsity of the conditional statement is not called into question. Indeed, subjects must suppose that the conditional statement is a rule that truly exists and applies to the cases at issue (if a rule may be violated, it must be assumed to exist). As Botterill and Carruthers (1999: 121) observe, “subjects need to engage in an extra level of processing in the indicative selection tasks, because in order to solve them they have to ask themselves: ‘Suppose this conditional applied. What would it rule out?’”. In the deontic versions of the task, instead, that is not necessary because subjects already know that the conditional rule applies. They have only to recognize whether it is being violated by examining the four cards. This is a fundamental difference between descriptive and deontic versions of the selection task. In particular, the indicative version presents a series of problems which the other task does not. Because of its complexity, in the next sections I will focus on the
indicative version of the task. Indeed, some questions arise as to it: what are subjects doing in the indicative version of the task? What kind of strategy do subjects employ? Are the most common responses of the subjects in this task as irrational as many scholars involved in the rationality debate have claimed?

3.3.1 Relevance-based explanation

Dan Sperber, Francesco Cara and Vittorio Girotto (1995) have proposed an explanation of subjects’ selections in the selection task based on a cognitive theory of linguistic comprehension, that is, Relevance Theory (Sperber & Wilson 1995). They have argued that the selection task does not call for reasoning or other similar activities, but is simply a task of selection guided by considerations of relevance. According to Relevance Theory, an utterance raises the hearer’s expectations of relevance, so that she tries to find an interpretation that may satisfy these expectations.7 The relevance of an utterance is the result of a trade-off between its cognitive effects, that is, the implications that can be drawn from it within the context wherein it is processed, and the processing effort required to derive those implications.

Let us consider how Sperber and his collaborators’ relevance-based explanation accounts for subjects’ performance on selection tasks with indicative conditionals. To begin with, they assume that in this version of the task subjects interpret the conditional rule as a universally quantified conditional statement, that is, $\forall x (Px \supset Qx)$. If understood in this way, the conditional rule is not directly testable. So, if subjects aim at evaluating the conditional rule, testable consequences must be inferred from it. At this point, Sperber and his collaborators argue that three main cases of interpretation of the conditional rule are possible (Sperber et al. 1995: 54-58). In the first case, the conditional is interpreted

7 The core of Relevance Theory is what Sperber and Wilson (1995: 17-19) call the Communicative Principle of Relevance, according to which “every utterance conveys a presumption of its own relevance”.
“biconditionally”, that is, as implying its converse ($\forall x(Qx \supset Px)$). In the second case, the conditional rule is interpreted as an existentially quantified conjunction, that is, $\exists x(Px \& Qx)$. These two interpretations lead respectively to selecting the $p$ card alone and the $p$ and $q$ cards. Although the two interpretations are logically inappropriate, they may be regarded as conversationally rational in ordinary discourse. Indeed, as Sperber and his collaborators (1995: 55) point out, a universally quantified conditional statement that does not have at least some instances will be considered as irrelevant in everyday conversation. In the third case, the conditional rule is interpreted either as implying a negative existentially quantified statement of the form not-$\exists x(Px \& \neg Qx)$ or as contradicting a positive existentially quantified statement of the form $\exists x(Px \& \neg Qx)$. These two interpretations lead both to the logically correct answer, that is, the selection of the $p$ and not-$q$ cards. Sperber and his collaborators (1995: 56) argue that, though these two interpretations are logically equivalent, “they are not computationally or representationally identical”. Accordingly, in order to elicit correct responses in the indicative version of the selection task, an experimenter should make subjects interpret the conditional statement as denying the existence of $p$-and-$\neg q$ cases. Clearly, because of its entailing two negations, this interpretation would be not so easily accessible in ordinary situations. According to Sperber and his collaborators (1995: 58-60), however, it is empirically demonstrable that if the experimenter manipulates the subjects’ expectations about cognitive effects and the effort required to achieve them by varying the information available in the experimental context and in particular the content of the conditional statement, the appropriate interpretation will be readily accessible to the subjects. As regards to the standard version of the task, Sperber and his collaborators (1995: 52) hold that “the artificiality of the task is so overwhelming as to discourage any but the lower expectations of relevance”. As a consequence, according to the relevance-based analysis, in such cases the subjects’ comprehension is guided by
considerations of least processing effort, leading to the selection of the \( p \) card alone or the \( p \) and \( q \) cards.

3.3.2 Probability and optimal data selection

Michael Oaksford and Nick Chater (1994; 1996) have attempted to account for subjects’ alleged normatively inappropriate responses in the indicative version of the selection task on the basis of what they call the Bayesian probabilistic approach to confirmation. According to their analysis, the selection task does not require deductive reasoning; rather, it calls for hypothesis testing involving optimal data selection. In their view, the subjects’ most common response can be seen as “optimizing the expected amount of information gained by turning each card” (Oaksford & Chater 1994: 609).

Let me begin with a clarification. Speaking of probability with reference to the selection task, one should assume that subjects will take the four cards of the task to be a sample from a larger set. Consider then a new version of the selection task in which subjects are asked to evaluate whether the usual indicative conditional (“if \( p \), then \( q \)”) is true or false by presenting, for example, 50 cards (instead of the usual four cards) in front of them. Suppose, moreover, that the number of cards displaying not-\( q \) is far greater than those with \( q \). Can we apply Popper’s hypothetico-deductive model to this version of the selection task? Will it give us any normative recommendation about how many cards we should examine in order to verify whether the conditional statement is true or false? The usual selection, checking the \( p \) and not-\( q \) cards, does not seem to be sufficient to test the conditional statement. Raymond Nickerson (1996) has compared this situation to the raven paradox (Hempel 1945): if subjects were asked to test the hypothesis “if it is a raven, then it is black”, they will probably consider ravens and examine whether they are black. In principle, there is another strategy that should equally increase their degree of confidence in
that hypothesis, that is, to consider non-black things and examine whether they are ravens or not. So, the paradox here is that any time you find a non-black non-raven (e.g., a white shoe), as well as a black raven, your degree of confidence in the truth of “All ravens are black” should increase. However, this seems to be puzzling. One of the well-known solutions of this paradox was proposed by the philosopher John Mackie (1963), who held that in such cases we should not aim at examining whether our evidence is relevant to increase our confidence in the hypothesis under examination, but, rather, whether it supports one of two possible alternative hypotheses, that is, one which states the independence of the property “raven” from the property “black” and the other which states its dependence. A similar line of reasoning can be assumed when discussing subjects’ selections in the indicative version of the selection task.

Michael Oaksford and Nick Chater have adopted an alternative model of hypothesis testing in determining the correctness of the subjects’ most common response in the selection task, that is, the Bayesian approach to optimal data selection. This model fits well with the solution proposed by Mackie with regard to the raven paradox. A fundamental assumption that Oaksford and Chater integrate to this model is what they call “the rarity assumption”, according to which people usually think that the probabilities of both $p$ and $q$ are low. That is, “the categories that function in everyday hypotheses about the world apply only to very small subsets of objects” (Oaksford & Chater 2003: 293). On the basis of this assumption and the Bayesian probabilistic model of confirmation, Oaksford and Chater have developed a framework for analyzing subjects’ responses to both standard and modified versions of the selection task. Indeed, according to their model, when faced with the selection task, subjects try to reduce their uncertainty about whether the conditional statement under examination is true by turning over the cards which maximise information gain, that is, the cards which maximally reduce their uncertainty about the truth of the
conditional statement. In doing so, they consider two alternative hypotheses: (i) that the conditional statement “if $p$, then $q$” is true (the consequent $[q]$ is dependent on the antecedent $[p]$) and that (ii) it is false (the antecedent $[p]$ and the consequent $[q]$ are independent). Just as a scientist may have two or more alternative hypotheses to examine and has to choose experiments which may provide the greatest “expected information gain” in order to decide among them, so too in the selection task subjects have to select the cards which are likely to give them the greatest expected information gain in order to decide about the two hypotheses mentioned above. So, when faced with the selection task, subjects would calculate the expected information gain of turning over each card, which, Oaksford and Chater explain, amounts to the difference between the “prior” uncertainty about the dependency hypothesis, that is, the uncertainty the subjects have before they have decided which cards to turn over, and the uncertainty about the same hypothesis after the card selections have been made. Supposing that subjects assume that $p$ and $q$ are rare (the rarity assumption), Oaksford and Chater calculate that the order of the expected information gain of turning over each card is:

$$
E(Ig(p)) > E(Ig(q)) > E(Ig(\text{not}-q)) > E(Ig(\text{not}-p))
$$

As to the standard version of the selection task, this order corresponds to the order of the subjects’ usual choices. As we already know, indeed, most subjects choose the $p$ and $q$-cards or the $p$-card alone. Where the rarity assumption does not hold, such as in some thematic versions of the task, the order of the amounts of the expected information gain just described changes. As Oaksford and Chater (1994: 610-614) have demonstrated, in such experimental situations subjects’ cards selections reflect that changed order.

Consequently, according to Oaksford and Chater, the choices of cards that have been taken to be evidence of human irrationality should be considered as due to a highly rational
reasoning strategy. In particular, this strategy fits well with our everyday situations where usually rarity is the rule, not the exception.

### 3.4 A two-step evaluation of subjects’ reasoning performances

Although they have proposed different explanatory models of subjects’ selections in the selection task, Oaksford and Chater and Sperber and his collaborators come to the same conclusion: namely, subjects do not choose the not-\( q \) card because they consider it as irrelevant with respect the problem they are faced with. On either line of argument, giving such a logically inappropriate response is rational. Accordingly, the most common response in the standard version of the selection task, usually interpreted as evidence of human irrationality, is reinterpreted as reflecting efficient cognitive strategies operating in a rational manner. Sperber and his collaborators (1995: 90) have maintained that subjects accused of providing normatively inappropriate responses in fact are giving pragmatically appropriate responses. On this view, subjects’ performances on the standard version of the selection task should be considered as rational, and this can be confirmed by looking at the pragmatics of the situation. In conversationalist terms, subjects should be regarded as conversationally rational (as is also claimed by Hilton and Politzer; see Chapter 5, Section 1.2). On the other side, Oaksford and Chater (1994: 609) have maintained that the selection task should be interpreted as a problem of inductive hypothesis-testing involving optimal data selection. Seeing it in this way, selecting the \( p \) and \( q \) cards becomes the most appropriate choice, caused by the use of a rational strategy in the field of inductive reasoning. Consequently, people who select the \( p \) and \( q \) cards in the standard version of the selection should be regarded as rational hypotheses testers.

Remember that in a previous section (see Chapter 5, Section 2) I have proposed a two-step normative framework, to be applied to the subjects’ task understanding and to their
responses. Such a framework assumes that one’s reasoning performance should be evaluated in terms of the legitimacy of her task interpretations and of the appropriateness of the reasoning strategy adopted. Interestingly, while providing different arguments about the normative appropriateness of the subjects’ responses, Relevance Theory and Oaksford and Chater’s approach focuses respectively on the subjects’ comprehension of the task and on the reasoning strategy they use. In particular, as seen above, their analyses seem to provide robust considerations in support of seeing the subjects’ understanding of the selection task and the reasoning strategies which they adopt as both legitimate. On this view, subjects are making rational selections. Looking closely at both the relevance-based explanation and the Bayesian analysis of the selection task, however, it seems to me there is something wrong with their explanations of the subjects’ most common response in the standard version of the selection task.

Let me begin with the relevance-based explanation. According to Sperber and his collaborators (1995), when faced with the selection task, subjects choose the cards guided by considerations of relevance, without employing any reasoning strategy. As seen above, given the fact that the standard version of the selection task gives rise to very low expectations of relevance in the subjects, Sperber and his collaborators have argued that the subjects’ most common responses are the best they can give in such a situation. The appropriateness of the subjects’ responses is evaluated on the basis of a hypothetical cost-benefit analysis of their cognitive economy, about which we cannot have any direct evidence. According to Sperber and his collaborators (1995: 62-89), however, this analysis is indirectly confirmed by a series of experiments they have devised. What they claim to have demonstrated in these experiments is that by changing the content and context of the conditional statement and the problem presentation in appropriate ways, the expected effects and the effort required to solve the task can be manipulated so as to prompt subjects
to provide either correct or incorrect responses according to the experimenter’s interests. While the standard version of the selection task discourages any but the lowest expectations of relevance in the subjects, some of the modified versions of the task give rise to strong expectations of relevance in them. In particular, such tasks make salient to the subjects that checking whether there are “p and not-q” cases, that is, counterexamples to the conditional rule, is more appropriate than checking whether there are “p and q” cases. For example, subjects were presented a story in which the leader of a secret religious sect, dubbed Haré Mantra, “was accused of having had some his sect’s virgin girls artificially inseminated” (Sperber et al. 1996: 63). While the head of the sect’s goal is to create an elite of “Virgin-Mothers”, he jokingly claims that “the women of his sect are, without exception, like any other women:

If a woman has a child [p], she has had sex [q]” (Sperber et al. 1996: 63).

Subjects are asked to imagine they are journalists who are trying to write an article on that sect. They are said that the women who are part of the sect have underwent a gynaecological survey, but the only evidence about the survey’s results that is accessible to them consists of four cards left on the gynaecologist’s desk, each recording, about a woman, whether she has children and whether she has had sex. These cards are half covered, so that in each, only part of the information is visible.

![Diagram of cards](image)

(p)  (not-p)  (q)  (not-q)
Subjects are told then that, while the doctor turns his back, they can take advantage by uncovering some of the cards. They are asked in particular to indicate (by circling them) the cards that should be uncovered in order to find out whether what the leader of the sect says (“If a woman has a child, she had had sex”) is true, as far as these four women are concerned, and to indicate “only those cards that it would be absolutely necessary to uncover” (Sperber et al. 1996: 63). In this modified version of the task, since checking whether there are virgin-mothers in the sect is more salient than checking whether there are normal mothers, focusing on whether the “p and not-q” case (a woman who has children and had no sex) occurs is easier than in the standard version. As is predicted by the authors’ relevance-based framework, the 75% percent of the subjects selected the p (“children: yes”) and not-q (“sex: no”) cards (Sperber et al. 1995: 62-66). The problem with this and other similar experiments devised by Sperber and his collaborators is that strong cues are provided in the experimental situation in order to prompt the subjects to select the correct counter-example. In some cases, the introductory story implicitly suggests the counter-example of the conditional rule which has to be tested. Making the task so easy, the interpretation of the resulting empirical data seems to be highly controversial. Indeed, other researchers have shown that these empirical results may be explained within other theoretical frameworks, which assume that people are really reasoning in the selection task and not only making a selection guided by considerations of relevance (see, e.g., Fiddick et al. 2000; Osman & Laming 2001). Going back to the interpretation of the standard version of the task, it seems to me that the empirical results presented by Sperber and his collaborators are not enough to support their interpretations of the subjects’ most common selection.

I now turn to Oaksford and Chater’s analysis. In order to explain the rationality of subjects’ selections in the standard version of the selection task, Oaksford and Chater make
some fundamental background assumptions about how the task is interpreted. Indeed, they assume that subjects both approach the task as a problem in inductive hypothesis-testing and assume that vowel and even numbers are relatively rare. According to these assumptions, the four cards presented in front of the subjects are taken to be as samples belonging to four general classes (a set of vowels, a set of consonants, etc.). Consider now an example by Oaksford and Chater (1994: 609). Imagine that a subject is evaluating the conditional statement “if you eat tripe [p], then you feel sick [q]” by examining four groups of subjects representing the four usual options: people who have eaten tripe (p), people who have not eaten tripe (not-p), people who are sick (q), people who are not sick (not-q). In order to verify whether the conditional statement is true or not, the subject would probably verify whether people who had eaten tripe (p) are sick. However, she can engage in another effective strategy, that is, to examine whether “people who are sick” (q) have eaten tripe. On the contrary, checking people who are not sick (not-q) in order to verify whether they have eaten tripe seems to be an unreasonable choice because this group of people will be too wide. So, Oaksford and Chater hold that the first two strategies will be more informative than the last one. And these are the strategies that Oaksford and Chater attribute to the subjects in the standard version of the selection task. As Stanovich (1999: 197) points out, however, they assume that in the standard version of the selection task subjects are thinking “in terms of sampling from classes of cards and have implicit hypotheses about the relative rarity of these classes”. In doing so, Oaksford and Chater are bringing details and information into the problem context, which are not present in the information as explicitly presented. Indeed, in the standard version of the task subjects do not receive any information about groups or classes of cards.

It is noteworthy that the 4% of the subjects select the p and not-q cards in the standard version of the selection task. This means that they have interpreted the task in the logically
correct way (and not in the way assumed by Oaksford and Chater). So, who is wrong? Subjects who assume that the four cards of the task are a sample from a larger set of cards or those who interpret the task correctly? Or both? According to my contextualist criterion, if the experimental context does not make explicit that the conditional rule is about only the four cards, both of these interpretations may be considered as correct. However, this is not the case: the conditional rule as explicitly stated is about only the four cards in the task. As a result, subjects who consider the selection task as an inductive hypothesis-testing task are wrong. As said above, they have been explicitly told that the conditional rule applies only to the four cards, not to some wider set of which they might be a sample. In other words, if subjects interpret the task the way Oaksford and Chater propose, their interpretation of the task is not legitimate. It does not match the information as explicitly presented in the experimental context. On this view, the difficulty for subjects is not in their reasoning badly, but in their misconceiving the situational context. They are using a rational strategy in the wrong situation. The upshot is that Oaksford and Chater’s analysis shows only that a certain percentage of subjects is rational according to their interpretation of the situational context. By assuming Oaksford and Chater’s analysis, we might maintain that, relative to their ordinary goals, the subjects’ most common response can be considered as rational, their real cognitive mistake resides in their representation of the task context which does not match the information as explicitly presented in the experimental context. Indeed, as seen above, most of the subjects tend to evaluate the indicative conditional with reference to a wider domain of evidence than the narrow one determined by the experimental context, as they do in ordinary life.
4. Concluding remarks

In this chapter, I have sketched out a context-sensitive consequentialist approach to rationality assessment. My approach provides a general normative framework which, I have maintained, is applicable both to subjects’ task interpretations and their responses. The two-step normative framework specifies two kinds of constraints that operate within the experimental context. On the one hand, the subjects’ understanding of the reasoning problem is always constrained by the information explicitly presented in the experimental situation. On the other hand, as is held by consequentialists, the evaluation of success in reasoning cannot be separated from the evaluation of success in achieving goals. Accordingly, the information explicitly presented in the experimental context places constraints on the goals and frames that it will be appropriate for the reasoner to activate when dealing with the task, which in turn determine the type of reasoning strategy that will be effective and efficient. In the last part of the chapter, I have applied it to the data on human reasoning from a well-known experimental reasoning task, that is, the Wason selection task. There is some ambiguity in the analyses of the subjects’ responses in the standard version of the task examined in this chapter: selections of $p$ and $q$ cards and, respectively, of $p$ and not-$q$ cards may be considered as appropriate depending on the reasoning problem’s interpretation that experimenters attribute to subjects. Such a way of approaching the subjects’ performances leaves the assessment of their rationality basically undecided, because there is no real agreement about which response is the correct one. In order to solve such an ambiguity, we need a more complex and refined normative background against which to assess the subjects’ reasoning performances than those usually assumed. As we have seen in this chapter, my two-step normative framework is able to solve the ambiguity by distinguishing the legitimacy of subjects’ task interpretations from the normative appropriateness of their responses. According to my framework, in order to
regard the subjects’ most common response in the standard version of the selection task as rational, we have to assume that subjects misinterpret the reasoning problem. Indeed, according to the researchers who have maintained that the most common response should be regarded as rational, subjects frame the reasoning problem in a wider context, that is, however, in a way which is inappropriate with respect to the situationally available information. What most subjects seem to do is to frame the problem in a way that makes sense of the situationally available information in the light of their ordinary goals. Within this ordinary frame, then, subjects tend to evaluate the indicative conditional with reference to a wider domain of evidence than the narrow one determined by the experimental context. But such an interpretation of the task is illegitimate because the problem presentation explicitly states that the conditional rule is about only the four cards in the task. So, according to my two-step normative framework, while subjects checking for the \( p \) and \( q \) cards rely on a rational reasoning strategy (at least in ordinary situations), their framing of the reasoning problem is illegitimate because it does not match the information available in the experimental context.
Chapter 6 – Concluding remarks

If it is the case that human reasoning is best evaluated within a context-sensitive consequentialist framework as I have argued in the previous chapter, what does this suggest for the study of human reasoning and rationality? I think there are implications for the way in which cognitive psychologists and philosophers conceive of some of the most important questions within the rationality debate. In particular, I would like to focus here on the main implications that the adoption of the proposed approach to rationality assessment would have in the analysis and the evaluation of subjects’ final responses to reasoning tasks.

As seen in the previous chapters, the researchers involved in the rationality debate can be divided into two camps on the basis of their commitment to either a deontological or a consequentialist conception of rationality. It is by relying on either of these conceptions that subjects’ performances on reasoning tasks have been assessed in terms of one’s ability either to comply with a set of normative principles (in the case of deontological approaches) or to reach a specific range of goals efficiently (in the case of consequentialist approaches). In any event, whether we accept a deontological approach to rationality assessment, either standard or non-standard, or opt for a consequentialist approach, such as those based on evolutionary or pragmatist considerations, we are still a long way from being able to provide a complete assessment of reasoning performances. On the one side, deontological accounts appear to be unsatisfactory, because by defining good reasoning as reasoning that conforms to a set of normative principles, they do not explain the normative force of these principles. On the other side, while agreeing with the general idea underlying the consequentialist approach to rationality assessment, I have maintained that
consequentialist approaches, whether based on evolutionary or pragmatist considerations, are flawed. With regard to consequentialist approaches of the evolutionary kind, I criticize the tendency of their supporters to justify the normative appropriateness of a reasoning performance by referring to the environment of evolutionary adaptation where its underlying cognitive mechanisms have evolved, rather than to its current context. Indeed, if reasoning is performed in a particular context and thereby is a situated event, it seems reasonable to think that it must be assessed with respect to that context, not with respect to the environment of evolutionary adaptation. As to the pragmatist kind of consequentialist approach to rationality assessment, by focusing exclusively on considerations concerning the value or significance of the goals pursued by the reasoners and the cognitive resources available to them for the attainment of their goals, pragmatists tend to neglect the distinction between the situational and cognitive constraints under which reasoners operate and consequently fail to distinguish the respective roles of those constraints in rationality assessment.

While the supporters of all of these four approaches have proposed different ways of characterizing what counts as rational behaviour, when it comes to the evaluation of reasoning performances they all display the same attitude, namely, they deal with the subjects’ task interpretation as a flexible object to be adjusted depending on their own research targets. On the one hand, those who trace subjects’ reasoning performances back to a predetermined normative standard assume that their subjects have understood the reasoning problem in the same way as they do and with the same background assumptions. On the other hand, other researchers hold that there is always an appropriate way to interpret the answers of the subjects as showing that they have understood the reasoning problem differently from how it was understood by the experimenter, so that their answers can be regarded as normatively appropriate to that which they respond to. In both these
cases, the subject’s task interpretation is not regarded as belonging to the reasoning performance and consequently as a part of what is subject to evaluation. So, researchers from both these trends leave it open to themselves to adapt their reading of the subjects’ task interpretation to their theoretical target (that is, to supporting a certain hypothesis about human rationality). Thus, any response can be regarded as normatively appropriate or inappropriate directly following from the task interpretation that researchers impute to the subjects. In contrast to this widely accepted attitude, in the fifth chapter I have argued that subjects’ task interpretations, as well as subjects’ responses to tasks, require a normative background against which to be assessed as to appropriateness or legitimacy. To address this question, I have proposed a two-step normative framework for situationally establishing the legitimacy of subjects’ task interpretations and the normative appropriateness of their responses. My approach specifies two kinds of constraints that operate within the experimental context. On the one hand, the subjects’ understanding of the reasoning problem is always constrained by the information available in the problem context. On the other hand, as is held by consequentialists, the evaluation of success in reasoning cannot be separated from the evaluation of success in achieving goals. With regard to the first set of constraints, I have observed that participants and experimenters have many different interpretations of reasoning problems. Regardless of how experimenters interpret a reasoning problem, if the subjects’ framing of the problem matches the contextual information available in the experimental situation, it should be regarded as a legitimate interpretation of that problem. On this view, the same problem may be framed in different ways on the condition that the resulting representation matches the information available in the problem context. As to the second set of constraints, depending on the goal-structure of framings, in order to assess reasoning performances we should first ask what purpose reasoner’s answers are aimed at and then consider whether they are
correct or incorrect with respect to it. Once a goal has been set, a reasoning performance is normatively appropriate depending on its effectiveness in achieving that goal. As a result, such an approach to rationality assessment holds that the information explicitly presented in the experimental context places constraints on the goals and frames that it will be appropriate for the reasoner to activate when dealing with the task, which in turn determine the types of reasoning strategy that will be effective and efficient.

It is clear that the subject’s response does not explicitly suggest how she has arrived at that response. We can only make suppositions about what has happened. However, the fact that a subject does not respond to the reasoning problem as the experimenter expects does not suffice to regard her response as normatively inappropriate. Nor can we hold that, given the subject’s task interpretation, her reasoning performance will be always normatively appropriate. If we assume that the context of evaluation is determined also by the reasoner’s own interpretation of the problem context, the appropriateness of a given reasoning performance would demand of the reasoner too little: it would demand of her that she behaves in accordance with her own interpretation of the situation but not that her understanding of the task displays a correct grasp of the situation. By relying on the two-step normative framework, we take into account both the legitimacy of the subject’s task interpretation and the normative appropriateness of her response. In conformity with such a framework, when analyzing the subjects’ final responses to a reasoning task, we are faced with four different possible cases: (i) the subject’s task interpretation is legitimate and her response is normatively appropriate with respect to that interpretation; (ii) the subject’s task interpretation is legitimate but her response is normatively inappropriate with respect to that interpretation; (iii) the subject’s task interpretation is illegitimate but her response is normatively appropriate to that interpretation; (iii) the subject’s task interpretation is
illegitimate and her response is normatively inappropriate anyway. Let us consider these four cases in greater detail.

With regard to case (i), two sub-cases may be distinguished. In the former, the subject’s response corresponds to what the experimenter regards as normatively appropriate under her interpretation of the task. So, also the subject’s task interpretation is assumed to correspond to the experimenter’s. Such a condition resembles that envisaged by evolutionary psychologists. According to them, every time we devise a reasoning problem in which the information available in the problem context is presented in an appropriate form and the problem itself recalls a problem from the environment of evolutionary adaptation, we prompt the subjects to frame the problem accordingly, which leads them to a normatively appropriate response. In such a case, on the basis of evolutionary considerations, we assume that both the subject’s task interpretation and her response match those assumed by the experimenter. In the latter sub-case of (i), the subject’s response diverges from that expected by the experimenter, but can nevertheless be regarded as normatively appropriate with respect to a legitimate task interpretation (different, in turn, from that assumed by the experimenter). Such a way of dealing with empirical data is open to two of the approaches to rationality assessment we have considered: the conversational approach and the consequentialist approach inspired by pragmatism. By adjusting the subject’s task interpretation or, respectively, the goal with respect to which the reasoning performance is expected to be effective, these approaches are able to re-interpret any reasoning performance as rational. Indeed, we have seen that the former places too weak constraints on the subject’ task interpretation, while the latter leads directly to relativism. If we rely on the notion of objective context at both steps of assessment, as the two-step normative framework proposes, several suggestions coming from these two trends of research can be accepted while avoiding their undesirable consequences.
As to case (ii), the experimenter assumes that the subjects have understood the problem presentation in the same way as she does and with the same background assumptions, so that she can assume that they have framed the reasoning problem in a legitimate way, but with respect to this interpretation of the problem, the experimenter assesses the subjects’ responses as normatively inappropriate. Such a case corresponds to the one in which an experimenter compares the subjects’ reasoning performance with a predetermined normative standard without taking into consideration what the subjects are really doing. Such a way of assessing the subjects’ responses to reasoning tasks resembles the attitude taken by the supporters of the Irrationality Thesis. On their view, people understand correctly the reasoning problem but, in solving it, rely on heuristics which do not lead to good outcomes reliably enough.

Turning to (iii), the case in which the subjects’ framing of the reasoning problem does not match the information available in the experimental context, but they may be held to use a rational reasoning strategy nevertheless, it is Oaksford and Chater who can be described as construing the experimental results in this way. In their view, contrary to the judgement of most psychologists, the most common response to the standard version of the selection task should be regarded as rational. In arguing for this claim, they have to assume that the subjects consider the selection task as a task of inductive hypothesis-testing. If it were so, however, the subjects would misinterpret the reasoning problem. Indeed, they have explicitly been told that the conditional rule applies only to the four cards, as opposed to some wider class of which these cards might be a sample. Oaksford and Chater hold that if we accept this task interpretation, we can admit that the subjects are using a rational reasoning strategy which leads most of the time to successful outcomes in ordinary life. In a case such that the one envisaged by Oaksford and Chater, I hold that the difficulty for subjects is not in their reasoning badly, but in a misperception of what the situational
context is. In other words, they are using a rational reasoning strategy in the wrong situation.

Finally, there is case (iv). In such a case, both the subject’s response and her interpretation of the task are inappropriate given the information available in the situational context. Even if the experimenter hypothetically assumes as appropriate the subject’s representation of the problem, she will conclude that the reasoning strategy the subject has used is inappropriate with respect to that interpretation of the problem. This case is, however, a mere combinatorial possibility. It seems to be very difficult to establish from the subject’s response, assuming that it is inappropriate to her task interpretation, what her task interpretation was like and whether it was itself inappropriate. Ideally, such a position may be assumed to hold by the supporters of a radical Irrationality thesis. They might argue that people have to change their cognitive strategies if they want to achieve their goals. These changes should be concerned with both people’s capacity to understand and frame reasoning problems and their ability to reason appropriately.

Coming to a conclusion, it seems to me that the supporters of the different approaches to rationality assessment that I have examined in the previous chapters have each a partial view of what could happen in a reasoning task. Although, as just seen, when assessing a reasoning performance, one is faced in principle with four different cases, they tend to consider only the one which fits best with their theoretical targets. They often do not care about the other possible cases. But, it seems to me, in assessing reasoning performances we are not allowed to reject any of the four cases a priori. As a consequence, the two-step normative framework presented here can be seen as more satisfactory, in that it enables us to conceive of all the four cases. Thus, the two-step normative framework of the context-sensitive consequentialist approach appears to have a potential for use in combination with traditional approaches such as the evolutionary and pragmatist ones in particular, towards
an overarching framework for the assessment of human rationality. Understanding human reasoning and rationality depends upon how we conceive of the reasoning performance and its relation not only to the cognitive context, but also to the situational context wherein it occurs. A satisfactory analysis of this relationship provides reasoning studies with a fundamental challenge that could require extensive research and development for some years to come.
REFERENCES


