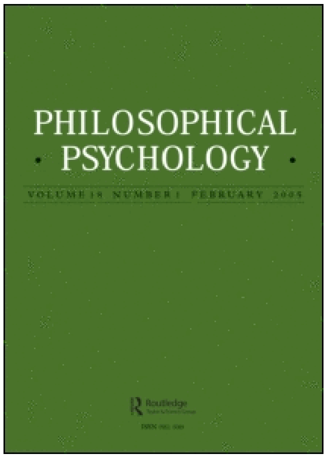

Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713441835

How to Play the Ultimatum Game: An Engineering Approach to Metanormativity

Benoit Hardy-Vallée; Paul Thagard

Online Publication Date: 01 April 2008
To cite this Article: Hardy-Vallée, Benoit and Thagard, Paul (2008) 'How to Play the Ultimatum Game: An Engineering Approach to Metanormativity', Philosophical Psychology, 21:2, 173 - 192
To link to this article: DOI: 10.1080/09515080801976581
URL: http://dx.doi.org/10.1080/09515080801976581

Philosophical Psychology
Vol. 21, No. 2, April 2008, pp. 173–192

Routledge
Taylor & Francis Group

How to Play the Ultimatum Game: An Engineering Approach to Metanormativity

Benoit Hardy-Vallée and Paul Thagard

The ultimatum game is a simple bargaining situation where the behavior of people frequently contradicts the optimal strategy according to classical game theory. Thus, according to many scholars, the commonly observed behavior should be considered irrational. We argue that this putative irrationality stems from a wrong conception of metanormativity (the study of norms about the establishment of norms). After discussing different metanormative conceptions, we defend a Quinean, naturalistic approach to the evaluation of norms. After reviewing empirical literature on the ultimatum game, we argue that the common behavior in the ultimatum game is rational and justified. We therefore suggest that the norms of economic rationality should be amended.

Keywords: Decision; Naturalism; Neuroeconomics; Normativity; Rationality

1. Introduction

Suppose you are offered $100 to share with someone else, but the catch is that the other person has to agree with how you propose to split the money. If they refuse your offer, you both get nothing. How much would you offer the other person? According to classical game theory, you should share as little as possible, and you should expect the other person to accept whatever you offer, because something is better than nothing. Perhaps, you should try to keep $99 for yourself and only offer $1 to the other person. In real-world situations, however, people often offer a 50–50 split, and recipients of smaller offers often reject them. Does this mean that

Benoit Hardy-Valle is a Postdoctoral Fellow in the Department of Philosophy, University of Toronto. Paul Thagard is a Professor in the Department of Philosophy, University of Waterloo. Correspondence to: Benoit Hardy-Vallée, Department of Philosophy, University of Toronto, 215 Huron Street, Toronto, ON, Canada, M5S 1A2. Email: ben.hardy.vallee@utoronto.ca

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/08/020173-20 © 2008 Taylor & Francis
DOI: 10.1080/09515080801976581

such people are irrational, or rather that there is something wrong with the standard economic norms? How *should* you play the ultimatum game?

This article will try to answer this question by arguing that recent findings in neuroscience, experimental economics and psychology shed light on the decision-making process and on the way rationality should be understood. First, we discuss *metanormativity*, which concerns norms for the establishment of norms, and develop a Quinean, naturalistic approach to the evaluation of norms. Next, we review behavioral, anthropological and neurological studies of the ultimatum game. We argue that these results are problematic for traditional economic rationality, not because subjects deviate from the norm, but because the proposed norm is inappropriate.

2. Metanormativity: The Standard Picture

Metanormativity is the normative evaluation of practical or theoretical norms. While norms indicate what should be done or thought, metanorms stipulate standards for normative systems (Rasmussen & Den Uyl, 2005). Asking what a normative system of ethics should seek is an instance of a metanormative question. Metanorms 'regulate the conditions under which [normative] conduct could take place' (Rasmussen & Den Uyl, 2005, p. 39). Philosophical discussions about the nature of norms, such as the existence of moral norms or the naturalization of epistemology, are metanormative.

A metanormative account states what norms should be: how (and if) they can be revised and justified. These questions arise for instance when psychologists notice substantial discrepancies between norms of rationality and patterns of inference. Empirical studies have shown that on many occasions normal subjects do not follow probability theory, formal logic or decision theory (Kahneman & Tversky, 2000). A metanormative discussion addresses the question whether or not these results justify revising normative standards.

According to what Edward Stein labeled the 'standard picture', these empirical studies have no normative significance. To be rational is 'to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth' (Stein, 1996, p. 4). Normative theories describe the nature of an ideal rational agent whose conduct conforms to the prescriptions of these theories. Mathematical proofs, such as consistency proofs or existence theorems, justify the normative validity of these ideal-agent theories. Thus, standards of rationality determine what agents should do or think, and why they are wrong when they do not follow these rules.

The standard picture implies that normative and descriptive projects are independent. Following Frege, normative enquiries should 'separate sharply the psychological from the logical, the subjective from the objective' (Frege, 1884/1980, p. x). Whatever we might learn about human psychology or brain functioning, it will not tell us what rational agents *should* do (Hempel, 1961). Facts are relevant to norms only in evaluation: it is mandatory to compare norms to facts in order to assess subjects' performance. Norms determine which facts should be observed if agents behave rationally, but the establishment and justification of norms is a separate project.

The standard picture is a metanormative account of theoretical and practical rationality. Peacocke's account of conceptual norms is an example of a standard picture of theoretical rationality (Peacocke, 1992). A normative account spells out the possession conditions of a particular concept. These conditions are specific inferences one must be disposed to draw if one masters a concept. For instance, one possesses the CONJUNCTION concept if and only if one is disposed to find these inferential transitions 'primitively compelling':

$$\frac{P \quad Q}{\therefore P \,\&\, Q} \qquad \frac{P \,\&\, Q}{\therefore P} \qquad \frac{P \,\&\, Q}{\therefore Q}$$

In Peacocke's account, the normative and the descriptive are sharply distinguished. On the one hand, logicians, philosophers or semanticists determine which rules are normatively correct. On the other hand, psychologists conduct experiments that study how subjects follow or fail to follow these patterns of inference. Johnson-Laird, for instance, studied logical reasoning and proposed a theory of mental modeling that explains discrepancies between logic and ordinary reasoning (Johnson-Laird, 1983). The psychological and the logical are, for these researchers, sharply separated.

The standard picture is also the received view of practical rationality. On the one hand, theoretical economics determines how rational agents should make decisions. Savage (1954) formalized the rules of individual decision-making, while Nash (1953) and Von Neumann and Morgenstern (1944) formalized the rule of strategic decision-making. These formal works set down in detail how preferences are organized (e.g., completeness, transitivity) and the rational outcome of games (equilibria). On the other hand, psychologists and behavioral economists study how people make decisions. In most of their experiments, Tversky and Kahneman (1981) asked subjects to choose among different options in order to assess the similarity between natural ways of thinking and normative decision theory. For instance:

> Imagine that the United States 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 estimates 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 600 people will be saved and a two-thirds probability that no people will be saved.

> Which of the two programs would you favor? (Tversky & Kahneman, 1981, p. 453)

Most of the respondents opted for A, the risk-averse solution. Respondents were offered the following version:

– If Program A is adopted, 400 people will die.

– If Program B is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. (Tversky & Kahneman, 1981, p. 453)

Although Program A has exactly the same outcome in both versions (400 people die, 200 will be saved), in the second version Program B is the most popular. Thus, not only are subjects risk-averse, but their risk-aversion depends on the framing of the situation. Subjects have a different attitude whether the same situation is presented as a gain or as a loss. Tversky and Kahneman (1986) concluded from their studies of human bounded rationality that the normative and descriptive accounts of decision-making are two separate projects:

> Deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system [...] the normative and the descriptive cannot be reconciled. (p. 272)

Thus, Tversky and Kahneman's research program adopted the standard picture of metanormativity. Instead of suggesting new norms for decision theory, they proposed more *descriptively* accurate models of practical rationality. According to their Prospect Theory, normal subjects tend to make certain errors because they draw upon cognitive heuristics and biases that mislead reasoning (Kahneman & Tversky, 1979). Hence, this position is compatible with the claim that norms of rationality cannot be amended, justified or established by empirical findings; these findings explain irrationality. These examples do not show that current normative frameworks are wrong, but rather that common wisdom assumes that complying with them will lead to optimal performance.

In sum, the standard picture consists of adopting the following metanormative views:

- S1. Norms are justified *a priori*, by their logical virtues (consistency, existence theorems).
- S2. Norms and facts belong to two different projects that cannot inform each other.
- S3. Norms are not to be revised once the axioms are laid down (unless another axiomatization is more logically elegant or wider in scope). We contend, however, that there should be a better agreement between normative and descriptive projects.

3. The Engineering Account

Norms of practical rationality, in the standard picture, are rules that an ideal agent should follow. The establishment, justification and revision of norms is a logical enterprise. Facts can only help us to assess or improve subjects' performance. Nevertheless, as we argue in this section, the standard metanormative account neglects one important dimension of practical rationality: the effectiveness of norms. We suggest that an engineering account of metanormativity is a better alternative.

In the standard picture, effectiveness is taken for granted but is not evaluated. That is to say, it is supposed that an agent that follows rational-choice theory prescriptions will be better off than an agent that deviates from rational standards, yet this assumption is never subject to any empirical verification. The quality of a decision is therefore, a function of the fit between the actual agent and the normative theory. However, compliance with normative theory may not lead to a better pay-off. Take the prisoner's dilemma for instance. In a *finitely* repeated prisoner's dilemma, the unique, dominant-strategy Nash equilibrium is defection in every game (Luce & Raiffa, 1957, pp. 97–102). Backward induction reasoning starts with the last game: since it is a one-shot prisoner's dilemma, agents cannot retaliate in a future game and thus should defect. Knowing that the last game will lead to common defection, players have no incentive to co-operate in the penultimate game. The same reasoning applies to the antepenultimate, the one before, etc., until the first game. Thus nobody will cooperate. This result however violates a common intuition: why not cooperate, at least at the beginning? Even Luce and Raiffa (1957, p. 100) recognized that they would not follow backward induction in the finitely repeated prisoner's dilemma. Moreover, all experimental studies indicate that subjects cooperate in prisoner's dilemma and will defect, in the *finitely* repeated version, only in the last game (Ledyard, 1995; Sally, 1995). According to the standard picture, this cooperation is not rational because it does not comply with game theory.

Against the standard picture, one can argue that the goodness of decisions is not only a matter of rule-following—accordance with normative standards—but also of effectiveness. Economic agents value different things (money, reputation, happiness, pleasure) and norms of economic rationality should not lead to inferior outcomes. Contrary to standards accounts, *consequentialist* accounts of metanormativity stress the importance of the outcomes of norms. Rational decision-making is considered as an 'effective means of achieving some goal or range of goals' (Samuels, Stich, & Faucher, 2004, p. 166). A normatively correct decision owes its correctness not to its compliance with an abstract normative theory, but to its effectiveness in achieving goals such as maximizing money, utility, fitness, or happiness. Rule-following is important when it contributes to goal achievement, but does not define rationality. According to Samuels et al., consequentialism about normativity implies that what is good reasoning or deciding may vary across contexts. The rules of right reason are not universal but must be relativized to agents and environments: if a procedure leads to desire-satisfaction for a certain kind of agent in a certain environment, then it is rational to use this procedure in this environment.

In this account, norms are justified by their effectiveness. If a procedure succeeds best in attaining a particular goal in a certain context, it is therefore, a normatively correct procedure in that context. Descriptive theories are highly important for norm consequentialism because they provide the list of possible procedures out of which the effective ones can be selected. It is factual knowledge that indicates which norms are rational, and it is also factual knowledge that will indicate which norms should be revised.

Consequentialism requires that norms are evaluated according to their effects. Goldman's reliabilism is a variety of epistemic consequentialism, which requires that reasoning processes are reliable, that is, that they tend to lead to true beliefs and to avoid false ones (Goldman, 1992). Good reasoning processes should be verific, that is, truth-aimed, and their outputs are justified if they are formed by a reliable psychological process. 'Epistemics', according to Goldman, is the study of those cognitive and social processes. The variety of consequentialism we advocate here is of practical, not epistemic, nature. It requires that norms help goal achievement in an effective fashion. Norm-building consists in designing effective and feasible procedures for decision-making, given the cost and benefits of the procedures and their competitors.

Our account of practical norms of rationality is naturalistic. Many criticisms of naturalistic projects highlight their incompatibility with normativity (Kim, 1988). Norms and facts belong to two different and incompatible styles of explanation. In the first, 'things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be', while in the second 'one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen' (McDowell, 1985, p. 389). Thinking that the first can be reduced to the second is 'bald naturalism' (McDowell, 1994). A naturalistic approach to normativity, however, is not committed to a reductionist conception of norms wherein one obtains an 'ought' from an 'is'.

When Quine (1986) replied to critics of naturalized epistemology, he did not propose to eliminate normativity, but suggested instead that a naturalistic conception of normativity is possible if one sees how it shares a deep similarity with engineering:

> 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. Like any technology, it makes free use of whatever scientific findings may suit its purpose. It draws upon mathematics in computing standard deviation and probable error and in scouting the gambler's fallacy. It draws upon experimental psychology in exposing perceptual illusions, and upon cognitive psychology in scouting wishful thinking. It draws upon neurology and physics, in a general way, in discounting testimony from occult or parapsychological sources. 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. (pp. 664–665)

If naturalized normative epistemology is the 'technology of truth seeking,' then naturalized practical rationality is the 'technology of decision-making'. It also 'makes free use of whatever scientific findings may suit its purpose'. If we want to establish, justify or evaluate a practical norm, different disciplines are required. It is important to note here that the relationship between engineering and normativity is an *analogy*, i.e., a partial identity. Similarities, analogy and metaphors play an important role in scientific discovery, rather than justification (Holyoak & Thagard, 1995, chap. 8). They are 'vital at the growing edges of science and philosophy'

(Quine, 1978, p. 161). We use them in contexts where there are no standard conceptions, solutions or rules: analogies are approximate guide for exploring unknown territories, not precise map of known territories. The naturalization of normativity is still uncharted territory where analogies can provide some guidance.

Since it is an analogy, the relationship between engineering and normativity should not be read literally, but interpreted as a heuristic for developing a naturalistic account of normativity. The analogy is not intended to suggest that only engineers deal with normative issues or that normativity is not addressed by other disciplines. It merely highlights two features of engineering that may be relevant for philosophers interested in a naturalistic account of normativity: the instrumental character of norms (given a particular goal, what is the best procedure for attaining it?) and the central importance of factual knowledge in designing effective norms.

Consequently, the value of a norm can be assessed from different perspectives: the emotional affect, monetary outcome, social utility, ethical rightness, computational tractability, psychological plausibility, or biological relevance of a norm can all be relevant for judging it. A naturalistic account of practical norms needs factual inputs from psychology, anthropology, artificial intelligence, and neuroscience. In contrast, standard rational-choice theory needs only existence theorems and consistency proofs. Of course, rational-choice theory may also play a role in norm-building, but as one of the tools required for designing effective norms rather than the main source of prescriptions.

Diverse disciplines will therefore, have a contribution to make to norm-building. Psychology analyzes the mechanisms humans use to make decision. Experimental research can reveal either if inference patterns comply with standard norms, or if they are effective in particular environments. If some inference patterns are defective, psychological studies may help correct them. Psychology can also set the limits of our bounded rationality by showing what procedures we tend to rely on. For instance, knowledge of the biases and shortcomings inherent in our prediction of future appraisals of choices, or 'affective forecasting', is important for devising norms of decision (Gilbert, 2006). Anthropological studies of cultural practices and social norms may be useful for assessing the value of a particular social strategy. In many game-theoretic situations, the pay-off of one's decisions is determined by the decisions of other players. Hence, in a culture that values reciprocity highly, it can be suboptimal to behave selfishly, while in another culture it might be desirable. Cultural norms can be at least as important as 'rationality postulates' for the development of locally optimal game-theoretic strategies. While psychological and anthropological studies can investigate whether norms of rationality are effective in attaining particular goals, artificial intelligence and computer science can contribute to norm evaluation by assessing the tractability and effectiveness of particular normative procedures. Some problems cannot be solved by 'brute force' algorithms and require simpler heuristics. Finally, neuroscience is a powerful tool for deciphering the motivations and mechanisms that guide decision-making (Hardy-Vallée, 2007). Neuroscientific measurements can relate hypotheses about location and activation of neural circuitry to cognitive mechanisms, economic parameters and observable behavior.

Imaging studies, for instance, could show why subjects have strong preferences for certain outcomes although their pay-off is equivalent. To make effective decisions, it is mandatory to know what human brains usually value.

Philosophical norm-building is thus closer to engineering than science. Science, from quantum mechanics to sociology, aims at describing adequately the physical, biological and social world. Engineering seeks the attainment of particular goals through feasible means: building a bridge, launching a rocket, securing a computer network, etc. (Koen, 2003). Engineers do not build bridges or launch rockets, but spell out what one *should* do in order to attain a particular goal. The work of the engineer consists not just in describing, but also in putting forth prescriptions: you should do A in order to attain B. The required normativity is not transcendental, *a priori* or absolute, but instrumental. This normative process will recruit different kinds of knowledge: economics, meteorology, physics and statistics may all be relevant for the construction of a particular bridge. Moreover, the engineer, while devising prescriptions for a particular task, must also follow other norms, for instance concerning construction (certain materials are not allowed), security or IEEE standards. Hence, the normative production of the engineer is itself subject to normative constraints.

There is no absolute division between science and engineering but the distinction is nonetheless important. Technology is based on scientific knowledge and science progresses, thanks to technology: building spacecraft requires physics, while microbiology requires computers and microscopes. The difference, as philosophers of technology usually acknowledge, is not semantic but pragmatic: science and engineering use the same knowledge but tend to focus on different goals (see Kroes, 1998 and Scharff & Dusek, 2003). The scientist's job is done when an explanation is provided, while the engineer's job is done when a plan, a design, or a procedure is worked out.

Thus, an engineering task might be construed as the setting up of prescriptive procedures that allow one to achieve a particular goal, given the cost and benefits of the procedures and their competitor. More precisely, an engineering task will consist mostly in constraint satisfaction (Thagard, 2000): economic, temporal, physical, ethical and epistemic imperatives constrain the engineer's work. The bridge must be safe, functional, resistant, not too expensive, designed according to the current knowledge of material and geography of the area, and all of that at the same time. As the formula says: 'Quick, Cheap, Good: Pick two'. Hence, the engineer must compromise, accord priority to certain problems and try to attain these often-incompatible objectives. In the end, all engineering is about constraint satisfaction (Koen, 2003, p. 10).

Again, we reiterate the importance of the metaphorical character of Quine's remark and our heuristic use of this metaphor. We do not want to argue for a sharp distinction between engineering and science, but rather acknowledge that, generally speaking, the former is ultimately concerned with achieving practical goals (construction, design, etc.), while the latter is ultimately concerned with theoretical goals (constructing models, theories, hypotheses, etc.). Perhaps, the better illustration is the distinction between how computer science and software engineering may deal with computational complexity. The computer scientist could say of a problem that it is not decidable in polynomial time, while the software engineering would develop a tractable approximation or work-around. Thus, the task here is not to strongly distinguish science from engineering (a futile goal), but to show that the engineer's concern with facts and prescription provides, as Quine noted, a convenient illustration that naturalistic philosophers are not limited to description. Critics of naturalistic projects in philosophy (e.g., Kim, 1988) pointed out that once we treat a philosophical problem from a naturalistic perspective, we effectively change the subject, because philosophical problems are in part prescriptive. Naturalistic epistemology according to Kim forgoes justification and therefore, fails to be epistemology since 'for epistemology to go out of the business of justification is for it to go out of business' (Kim, 1988, p. 391). Yet Quine's engineering analogy, as we try to develop it, suggests that naturalistic philosophy can be normative. While critics of naturalism stress that descriptive knowledge such as scientific knowledge fails to provide a basis for normative claims, we stress that naturalistic normativity is possible if normative philosophy is compared with engineering rather than science. Note that this account is agnostic as to what the ultimate goals of epistemology, practical reason, ethics, philosophy or science should be: like Quine, we direct our reflection on the normativity of means rather than the normativity of ends.

Therefore, the engineering account of normativity consists of the following metanormative views:

M1. Norms are justified by their excellence in achieving particular goals, given their cost and benefits.
M2. Descriptive theories are useful to devise and assess norms.
M3. Norms can be revised by any empirical findings showing that other procedures are more effective.

In a later section, we will use empirical studies of the ultimatum game to show how norms of practical rationality can be revised in the light of empirical results, according to the generic account outlined here. Hence, we shall argue, you rationally *should* make a fair offer in the ultimatum game. First, we need to clarify how rationality should be assessed.

4. Assessing Rationality

It is important to distinguish internal and external assessment of rationality. An 'internal' (or subjective) assessment of rationality is an evaluation of the coherence of intentions, actions and plans. Internally rational actions make sense 'from the point of the cognitive and conative perspective of the agent' (Stueber, 2006, p. 49). An 'external' (or objective) assessment of rationality is an evaluation of the effectiveness of a rule or procedure. It assesses the optimality of a rule for achieving a certain goal. Externally rational actions make sense 'relative to a given set of environmental parameter that include the agent's desire but not his belief'

(Bermúdez, 2002, p. 260). An action can be rational from the first perspective but not from the second one, and vice versa. Hence subjects' poor performance in probabilistic reasoning can be internally rational (subjects may have good reason to choose a certain prospect) without being externally rational (their behavior is still suboptimal). The Gambler's fallacy is and always will be a fallacy: it is possible, however, that fallacious reasoners follow rational rules, maximizing an unorthodox utility function. This distinction thus specifies two nonexclusive ways in which someone can behave irrationally. One can be *externally* irrational if the outcome of an action is suboptimal; in this case the attribution of irrationality requires data about the agent and the outcome. One can also be *internally* irrational if, regardless of the outcomes of the action, the agent's desires and action performed are incoherent. Akrasia (acting against one's best judgment) is a form of practical irrationality because a desire (I want to stop smoking) and the action (lighting up a cigarette) are not coherent.

A growing number of studies in perception, psychology and psychophysics employ statistical decision theory in a contextualized fashion (see Ernst & Bulthoff, 2004; Geisler & Kersten, 2002; Maloney, 2002 for review). Many studies show that common patterns of perception and action are close to normative standards: speech intelligibility (Müsch & Buus, 2001), visual cue combination, (Landy & Kojima, 2001) motion perception (Weiss, Simoncelli, & Adelson, 2002), intersensory interactions (Gepshtein & Banks, 2003), action selection (Trommershäuser, Maloney, & Landy, 2003a, b) and sensorimotor control (Kording & Wolpert, 2006). These researchers assume that many constraints affect subject performance; for example, Trommershäuser et al.'s (2003a) model of movement planning integrates joint mobility, muscle tension changes, rate of change of acceleration and torque change in a biomechanical cost function. Assumptions are also made about the statistical structure of the environment: Weiss et al.'s (2002) model of motion perception stipulates that low velocities are more probable than high velocities. Another common assumption is that perception is heavily noisy and thus always takes place under uncertainty.

Once all these constraints and assumptions are integrated, it is possible to derive normative predictions about rational perception or action selection. Human performance can then be compared to normative predictions. When predictions and performance are on par, researchers can deduce that they modeled adequately the task and its constraints. Studying optical illusions, for instance, Weiss et al., conclude for that they are 'the best solution of a rational system designed to operate in the presence of uncertainty' (p. 598). Geisler and Kersten (2002) explain that Weiss et al.'s assumptions about the probability and likelihood distributions of velocities are 'incorporated into the visual system arise through a combination of evolution and perceptual learning' (p. 509).

This is where the distinction between internal and external assessment of rationality is important: research on perception does not prove that optical illusions are not illusions. They are still externally irrational but appear as internally rational, that is, produced by a rational Bayesian mechanism. It is still wrong—externally irrational—to infer that two lines are of different length in the Müller–Lyer illusion even for someone well-versed in the psychology and psychophysics of perception because, when we measure these two lines, they are of different length. The external rationality is always assessed relative to an external measure (something in the environment) while internal rationality requires only data about the agent (its neural/cognitive mechanisms). Perception researchers compare different normative models and select the one that accounts for human performance (Landy & Kojima, 2001).

The moral of this story is that we should expect an overlap between normative and descriptive theories, and the existence of this overlap is warranted by the hypothesis that agents adapt to their environment through natural selection and learning. Evolution and learning do not produce perfect agents, but ones that achieve some proficiency in perception, decision and action. Hence, on the normative side, we should ask what procedures agents should follow in order to make effective decisions given their environment and the constraints they face. On the descriptive side, we must assess whether a procedure succeeds in achieving goals or, conversely, what goals could a procedure aim at achieving. If there is no overlap between norms and facts, then either norms should be reconceptualized or facts should be scrutinized: it might be the case that norms are unrealistic or that we did not identify the right goal or value. Consequently, if humans systematically fail to solve a class of problems, we can consider whether subjects' performance is modulated by many constraints so that their behavioral output is a satisfying solution to a more complex problem, or to a similar problem in a different context. It can still be internally rational. The external rationality of the behavior is a different, but related, question.

Our conception of norms is similar to the Bayesian framework for perception and action. However, we are not only interested in finding the model that accounts for human performance, we also want to find the normative models that will yield the best result. Hence, we look for models that are both internally and externally rational. It is only in this situation that one can recommend a particular course of action and say 'you should do X'. Note that in certain models, there is a prescriptive dimension: for instance, from Trommershäuser et al.'s (2003a) study of movement planning and selection, one can infer a prescription of the optimal move. Although, this is not exactly a prescription for action because it does not tell us what we should do in economic decision-making (i.e., cooperate or not, make a fair or unfair offer), it embodies the main features of the engineering account developed here. This approach starts from a goal, then makes, as Quine (1986) says 'free use of whatever scientific findings may suit its purpose' (p. 664) (in this case, psychophysical data and statistical models) and finally derives an optimal policy. Thus, the approach represents a departure from a strict use of normative models of rationality—the standard picture—because it integrates these models with other empirical knowledge. However, experiments in psychophysics yield, in the end, a better *description* of human motor planning rather than policies—*prescriptions*—for optimal choices in economic contexts.

5. The Engineering Account and the Ultimatum Game

The ultimatum game (Güth, Schmittberger, & Schwarze, 1982) is a one-shot bargaining game between two players; call them the 'proposer' Alice and the 'responder' Bob. Alice offers an amount F (>0) of a total sum of money M to Bob; Bob may either accept or reject the proposal. If Bob accepts, he keeps F while Alice keeps the difference (M − F). If he rejects it, both players get nothing. According to game theory, rational agents must behave as follows: since Alice takes Bob to be a rational agent for whom any amount of money has a positive utility, she anticipates that he will accept any offer >0; she therefore should offer the smallest amount possible, in order to keep as much money as possible, and Bob should accept any proposed amount, because a small amount should be better than nothing. We will call this strategy for Alice GAME.

The ultimatum game has been studied in many contexts where different parameters of the game were modified: age, sex, the amount of money, the degree of anonymity, the length of the game, and so on (Oosterbeek & van de Kuilen, 2004; Samuelson, 2005). The results show a robust tendency: GAME is rarely played, because people tend to both anticipate and make 'fair' offers. When proposers offer about 50% of M, responders tend to accept these offers while rejecting most of the 'unfair' offers (<20% of M). We will call this strategy FAIR.

FAIR is almost universal. Henrich et al. (2005) found cultural variation in 15 different small-scale societies, but these variations exhibit a common pattern: proposers make offers that responders will consider fair, based on what people in this culture consider fair. In North American and European societies, an offer is considered fair when the responders receive at least 20% of the total amount (most of these offers, in fact, are close to 50–50). To the contrary, Machiguengas in Peruvian Amazon judge that low offers are worth accepting (Henrich, 2000). They believed that a little money is better than nothing, but did not feel 'cheated' by the proposers (Henrich, 2000, p. 977). Another study with Gypsy people in a neighborhood of Madrid indicates how fairness may be culturally determined: although 97% of the proposers offered an equal split, most of those who were offered nothing accepted the split, and justified their acceptance by saying 'if he really needs it . . .' (Pablo, Ramón, & Almudena, 2006). They thought it was fair to split this way. Machiguengas and Gypsies do not have the same understanding of a 'fair split' as Americans or Germans, but consider that fairness is important.

There are, of course, exceptions. People are closer to the game-theoretic strategy when players compete against a computer, when players are groups, when players are autistic people, and when players have been trained in decision and game theory, like economists and their students (Frank, Gilovich, & Regan, 1993; Hill & Sally, 2002; Robert & Carnevale, 1997; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). We will sketch later in this section an explanation for this pattern.

Why do normal, flesh-and-bone individuals tend to play FAIR instead of GAME? Given the simplicity of the game, it is rather improbable that subjects did not understand the rules or failed to compute the optimal strategy. Even when M is higher (e.g., $100) the results are similar: subjects are ready to give up a $30 offer because of its unfairness (Hoffman, McCabe, & Smith, 1996). One possible explanation is that people are generally disposed to be fair. Other explanations are also plausible: people may want to look fair in public, may fear punishment, or may worry about the experimenter's judgment. To help constrain hypotheses about why FAIR is preferred, we can look at the neurological underpinnings of decision-making. Research in decision neuroscience suggests that the best explanation for FAIR may be that people are emotionally disposed towards it.

Brain scans of people making ultimatum-game decisions indicate that three brain areas are specifically activated: the anterior insula, associated with negative emotional states like disgust or anger; the dorsolateral prefrontal cortex, associated with cognitive control, attention, and goal maintenance; and the anterior cingulate cortex, associated with cognitive conflict, motivation, error detection and emotional modulation (Sanfey et al., 2003). An unfair offer triggers, an aversive feeling in the responders' brain: the anterior insula is more active when unfair offers are proposed. This affective reaction is not a response to an unsatisfactory monetary reward, since the activation is significantly lower when the proposer is a computer. Thus, subjects feel repelled by the intentions and behavior of *human* proposer who offered less than the responders' minimally acceptable amount. They have a visceral reaction to another human's unfairness, as other experiments showed: offers and their rejection are associated with greater skin conductance (van 't Wout, Kahn, Sanfey, & Aleman, 2006). Moreover, the anterior insula activation is correlated with the degree of unfairness and with the decision to reject unfair offers (Sanfey et al., 2003, p. 1756). Unfair offers, therefore, elicit strong emotional response. Consequently, while behavioral results show how people react to unfair offers, neuroscientific results help understand why: humans, who value social relations, reciprocity, equity and fairness, suffer from perception of unfairness and thus are deeply motivated to punish cheaters. Other studies in decision neuroscience indicate that in similar games where cooperation is common but unexpected by game theory, players are ready to lose money for punishing untrustworthy players. Punishing free-riders in a public-good game—a game where players may contribute an initial endowment to a common pool or free-ride by contributing nothing and receiving their share of the common pool—activates the nucleus accumbens, a subcortical structure involved in pleasure (de Quervain et al., 2004). Moreover, cooperation in the prisoner's dilemma also has rewarding effects. Players who initiate and players who experience mutual cooperation display activation in nucleus accumbens and other reward-related areas (Rilling et al., 2002). Subjects in experimental games thus seem to follow a 'strongly reciprocal' strategy: cooperate spontaneously, cooperate with cooperators, do not cooperate with noncooperators and punish cheaters or free-riders, even at a cost to themselves (Fehr & Fischbacher, 2003; Gintis, 2000). Hence, even in one-shot blinded ultimatum games, where tit-for-tat reciprocity is impossible, strong reciprocity motivates players to behave fairly.

While the standard *Homo economicus* model represents agents as exclusively motivated by their material self-interest, economic theories of fairness put forth the picture of *Homo reciprocans*, an agent whose utility function incorporates social parameters (Bowles & Gintis, 2002; see Fehr & Schmidt, 2003, for a review). Economic theories of fairness fall into two categories: outcome-based models and intention-based models. The former explains fairness as the product of players' aversion to inequity (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). Players are sensible to the distributive consequences of strategic interactions and prefer resources allocations that reduce inequity: they negatively value a discrepancy between their own pay-off and an equitable pay-off (whether it is the mean pay-off or another player's pay-off). The latter explains fairness as the product of players' reciprocation of perceived kindness or unkindness (Dufwenberg & Kirchsteiger, 2004; Rabin, 1993). More than the outcome of an interaction, fairness is motivated by the attributed intention. For instance, in an ultimatum where the proposer's behavior is restricted to two options (50–50 and 80–20 split), the second option is the most rejected; when the proposer's options are 10–0 and 80–20, however, the second option is now accepted (Falk, Fehr & Fischbacher, 2003). Decision-makers value differently the same option whether it is perceived as an intention to be fair (valued positively) or not (negatively). Since both parameters appear to be important, many models integrate both intentions and outcomes (Falk & Fischbacher, 2006; Fehr & Schmidt, 2003).

A common feature of these models is the preservation of the optimality assumption: although they all suggest that the standard utility function should incorporate different parameters, they do not reject the idea that agents are internally rational: they maximize a nonclassical utility function. As the Bayesian perception framework discussed above suggested, normative models of decision-making can account for human performance once all relevant parameters are considered, such as the aversion to inequity and a motivation to reciprocate. In this case, economic models of other-regarding preferences account for—or more exactly, *rationalize*—actions that models of self-regarding preferences would not, such as the rejection of low offers in ultimatum games and costly punishment of free-riders in the public-good game. These rationalizations show the internal rationality of agents, that is, the coherence between their desires (preference for fairness) and actions (making fair offers). It is a different but related question to ask if the outcome of actions, regardless of the coherence of actions and desires, is optimal, since in this case the assessment of rationality involves information about the performance of the actions.

Another type of model may also account for human performance. In these evolutionary game-theoretic models, fairness in the ultimatum emerges as an evolutionary stable strategy (Huck & Oechssler, 1999). The idea is that if individuals interact in small groups, a preference for the rejection of low offers will come out of the replicator dynamics. Fearing that their offer might be rejected, proposers will therefore offer fair splits. Again, fair behavior may be rational, but this time there is no need to modify the standard utility function. Hence, different models may account for fair behavior without rejecting the optimality assumptions. It is still an open question which model is best, but a growing number of researchers in experimental economics and theoretical biology favor models of other-regarding preferences (Bowles & Gintis, 2004; Fehr & Fischbacher, 2003; Gintis, 2000).

However, as discussed above, we want to do more than account for human behavior. Whereas economic theories of fairness show the internal rationality of pro-social behavior, they do not prescribe or recommend courses of action. Our metanormative framework requires also an assessment of the *external* rationality of a performance. The important question is thus: Is it still correct to assert that the best move (according to external rationality) in an ultimatum bargaining with a normal human partner is to offer the smallest amount possible? *Should* you play GAME? We can see that in the ultimatum game, the standard norm—that is, the optimal move under the standard interpretation of the game—is not effective. The standard picture can accommodate cases where proposers anticipate that responders will reject offers under a certain amount: if proposers possess that information, than the optimal strategy is to make a fair offer. In this case, there is no need to introduce a novel norm. However, the standard picture fails to capture the effectiveness of the responder's rejection of low offers: responders accepting an unfair split will experience an aversive feeling, and would be happier in punishing unfair proposers than in accepting the offer, as showed by neural data. It is therefore, a better move to reject a few dollars.

We therefore suggest that, according to the engineering account of normativity, the normatively correct choice in the ultimatum game played against normal human subjects is FAIR. Responders should reject unfair offers and accept fair ones, either punishing an unfair proposer or receiving a fair part of the split. Since responders will systematically reject unfair offers, proposer should make fair offers; they will get 50% of the amount instead of 0% (that percentage may vary with the cultural determinant of what 'fair' means). It is important to note here that we argued that FAIR is a better strategy not because it is common or because it can be accounted for by alternative theories of preferences, but because it is effective. It is, therefore, both internally and externally rational.

Why is it that groups, autistic people, people playing against computers and subjects trained in rational-choice theory tend to play GAME instead of FAIR? For people trained in rational-choice theory, the answer is obvious, since they apply economic rationality principles to the game. We suggest that groups and autistic people deviate from what we consider the norm (FAIR) because they cannot use the same emotional-empathic mechanisms as normal human individuals. A group or an institution cannot have emotions, although its members can, and autistic people tend to have a limited competence in attribution of mental states (Baron-Cohen & Belmonte, 2005). Finally, since people do not take computers to have emotions and empathy, they do not feel bad in making unfair offers and do not anticipate the computer to experience moral disgust. But for normal subjects playing against other normal subjects, FAIR is a robust and effective norm. It is also effective in other domains. In management, studies show that fair organizations generate more profits, incur less legal pursuit and stimulate innovation (Brockner, 2006). Cross-national studies also reveal a strong correlation between a fair allocation of rights and resources (societal equality, democracy and human rights) and subjective well-being

(E. Diener, M. Diener, & C. Diener, 1995). In other words, fair countries tend to produce happy citizens. Fairness is thus not just good *per se*, but also when the consequences of fair policies are taken into account.

The ultimatum game is just one example of how an engineering account of practical norms should proceed. First, determine the objectives of the players, the constraints on their behavior, and the relevant data. Then try to find normative models that render human behavior internally rational. After that, select the strategy that will lead to optimal consequences, showing its external rationality. Finally, derive a policy or a prescription that states what one should do.

In the ultimatum game, the best strategy according to our account is playing FAIR, but another procedure may be rational in a different game. Our engineering account of rationality can be applied to other domains, since it consists of taking into account how humans value things like money, fairness, reciprocity, etc. We could apply our framework to any decision or game-theoretic problem, provided that we have enough neural and behavioral data. For instance, cooperation in the prisoner's dilemma can be expected and is rewarding for both players, as neural and behavioral studies suggest. Therefore, according to our account, it is rational to prescribe cooperation in the prisoner's dilemma—again, not because it is frequent, but because it has desirable consequences for both players. Cooperation helps to avoid the tragedy of the commons, the pattern of collective behavior where self-interested agents behaving selfishly create a collectively undesirable condition (Hardin, 1968). Cooperation is desirable and can be deemed rational not only because it can be rationalized (internally rational), but because the consequences of cooperation, in an environment of cooperators, is optimal (external rationality).

In all such cases, our engineering framework recommends that normative statements be evaluated as empirical hypotheses and that the effectiveness of strategies be justified by their expected consequences, not by their logical virtues. Consequently, this framework can draw a distinction between practices that are effective and those that are not, and between practices that are internally *versus* externally rational. Many strategies can be deemed internally rational once aversion to risk, uncertainty or inequity, is taken into account. Not all of them, however, will be *externally* rational. For instance, subjects may have good reasons to prefer saving, instead of killing, people, but if the consequences are the same, then our account agrees with the standard picture. We do not automatically reject standard models of rationality such as rational-choice theory, but require an empirical evaluation of any claim that 'X is the rational strategy'. More generally, our aim here is not to reject the standard picture *en bloc*, but to propose a consequentialist metanormative picture that may encompass it.

6. Conclusion

We have defended a consequentialist account of norms, and argued that normative projects can be construed, as Quine's analogy suggested, as engineering tasks. In the engineering account developed here, we show how a naturalistic conception of norms can be defended, and how it can be used to establish and revise norms of rationality. Norms are technologies evaluated on their effectiveness and thus should satisfy pragmatic constraints. According to the standard picture, agents are rational if they follow rational-choice theory. Contrary to the standard picture, our engineering account stipulates that norms should be empirically evaluated and that agents can be rational despite failure to comply with rational-choice theory. There is no *a priori* reason why deviating from classical norms should lead to a worse outcome than following them. Once we take into account how human valuations usually work, we can make normative recommendations that conflict with traditional game theory. However, we do not imply that probability theory or other normative frameworks are not relevant for decisions where information is perfect and complete, as in the Disease problem discussed above.

With the ultimatum game example, we showed why traditional norms of rationality were not adequate in dealing with the clash between norms and facts. While rational-choice theory recommends that proposers offer unfair splits and that responders accept any split, empirical studies reveal that normal subjects prefer proposing and accepting fair splits. This pattern is explained by the emotional mechanisms of subjects who evaluate splits. Neural studies indicated that responders have a negative emotional reaction to unfair splits. Proposers, because they anticipate such emotional reactions in responders, prefer to make fair offers. If we consider the monetary outcome and the hedonic impact of the FAIR strategy, we see that it is justified by the data. Fairness *is* rational.

Acknowledgements

This research has been funded by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. The authors are also thankful to anonymous reviewers and to Lutz Busch, Daniel Hausman, Angela Stanton and Chris Eliasmith for their comments.

References

Baron-Cohen, S., & Belmonte, M. K. (2005). Autism: A window onto the development of the social and the analytic brain. *Annual Review of Neuroscience*, *28*(1), 109–126.

Bermúdez, J. L. (2002). Rationality without language. In J. L. Bermúdez & A. Millar (Eds.), *Reason and nature*: *Essays in the theory of rationality* (pp. 233–264). Oxford: Oxford University Press.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of equity, reciprocity, and competition. *The American Economic Review*, *90*(1), 166–193.

Bowles, S., & Gintis, H. (2002). Behavioural science: Homo reciprocans. *Nature*, *415*(6868), 125–128.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, *65*(1), 17–28.

Brockner, J. (2006). Why it's so hard to be fair. *Harvard Business Review*, *84*(3), 122–129.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*(5688), 1254–1258.

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology*, *69*(5), 851–864.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, *41*(1), 20–26.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2003). Theories of fairness and reciprocity—Evidence and economic applications. In M. Dewatripont, L. Hansen, & S. Turnovsky (Eds.), *Advances in economic theory, Eighth World Congress of the Econometric Society* (pp. 208–257). Cambridge: Cambridge University Press.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation?. *The Journal of Economic Perspectives*, *7*(2), 159–171.

Frege, G. (1884/1980). *The foundations of arithmetic*: *A logico-mathematical enquiry into the concept of number* (2nd rev. ed.). Evanston, IL: Northwestern University Press.

Geisler, W. S., & Kerten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience*, *5*(6), 508–510.

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*(6), 483–488.

Gilbert, D. T. (2006). *Stumbling on happiness* (1st ed.). New York: A.A. Knopf.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*(2), 169–179.

Goldman, A. I. (1992). *Liaisons: Philosophy meets the cognitive and social sciences*. Cambridge, MA: MIT Press.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*(4), 367–388.

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248.

Hardy-Vallée, B. (2007). Decision-making: A neuroeconomic perspective. *Philosophy Compass*, *2*(6), 939–953.

Hempel, C. G. (1961). Rational action. *Proceedings and Addresses of the American Philosophical Association*, 35, 5–23.

Henrich, J. (2000). Does culture matter in economic behavior? Ultimatum game bargaining among the machiguenga of the peruvian amazon. *The American Economic Review*, *90*(4), 973–979.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2005). ''Economic man'' In cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Science*, 28(6), 795–815; Discussion 815–855.

Hill, E., & Sally, D. (2002). *Dilemmas and bargains: Theory of mind, cooperation and fairness*. Available at SSRN: http://ssrn.com/abstract=407040 or DOI: 10.2139/ssrn.407040.

Hoffman, E., McCabe, K. A., & Smith, V. L. (1996). On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, *25*(3), 289–301.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Huck, S., & Oechssler, J. (1999). The indirect evolutionary approach to explaining fair allocations. *Games and Economic Behavior*, *28*, 13–24.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.

Kim, J. (1988). What is naturalized epistemology? In J. E. Tomberlin (Ed.), *Philosophical perspectives* (vol. 2, pp. 381–406). Asascadero, CA: Ridgeview Publishing Co.

Koen, B. V. (2003). *Discussion of the method: Conducting the engineer's approach to problem solving*. New York: Oxford University Press.

Kroes, P. (1998). Technology, philosophy of. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, Retrieved December 12, 2007, from http://www.rep.routledge.com/article/Q103SECT3. London: Routledge.

Kording, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *18*(9), 2307–2320.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 111–194). Princeton: Princeton University Press.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York: Wiley.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world: Psychological and philosophical issues in perception* (pp. 145–189). New York: Wiley.

McDowell, J. (1985). Functionalism and anomalous monism. In E. LePore & B. McLaughlin (Eds.), *Actions and events: Essays on the philosophy of Donald Davidson*. Oxford: Blackwell.

McDowell, J. (1994). *Mind and world*. Cambridge, MA: Harvard University Press.

Müsch, H., & Buus, S. (2001). Using statistical decision theory to predict speech intelligibility. I. Model structure. *The Journal of the Acoustical Society of America*, *109*, 2896–2909.

Nash, J. F. (1953). Two-person cooperative games. *Econometrica*, *21*(1), 128–140.

Oosterbeek, H., Sloof, R., & van de Kuilen, G. (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics, 7*(2), 171–188.

Pablo, B.-G., Ramón, C.-R., & Almudena, D. (2006). "Si él lo necesita": Gypsy fairness in Vallecas. *Experimental Economics*, *V9*(3), 253–264.

Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.

Quine, W. V. (1978). A postscript on metaphor. *Critical Inquiry*, *5*(1), 161–162.

Quine, W. V. O. (1986). Reply to Morton White. In L. E. Hahn & P. A. Schilpp (Eds.), *The philosophy of W. V. Quine* (pp. 663–665). La Salle, MA: Open Court.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, *83*(5), 1281–1302.

Rasmussen, D. B., & Den Uyl, D. J. (2005). *Norms of liberty: A perfectionist basis for non-perfectionist politics*. University Park: Pennsylvania State University Press.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405.

Robert, C., & Carnevale, P. J. (1997). Group choice in ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *72*(2), 256–279.

Sally, D. (1995). Conversations and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*, 58–92.

Samuels, R., Stich, S., & Faucher, L. (2004). Reason and rationality. In I. Niiniluoto, M. Sintonen, & J. Wolenski (Eds.), *Handbook of epistemology* (pp. 131–179). Dordrecht: Kluwer Academic Publisher.

Samuelson, L. (2005). Economic theory and experimental economics. *Journal of Economic Literature*, *43*, 65–107.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Scharff, R. C., & Dusek, V. (2003). *Philosophy of technology: The technological condition. An anthology*. Malden, MA: Blackwell Publishers.

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Oxford University Press.

Stueber, K. R. (2006). *Rediscovering empathy: Agency, folk psychology, and the human sciences*. Cambridge, MA: MIT Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003a). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision*, *16*, 255–275.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003b). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *20*(7), 1419–1433.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and psychology of choice. *Science*, *211*, 453–458.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(4), 251–278.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the ultimatum game. *Experimental Brain Research*, *169*(4), 564–568.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (3rd ed.). Princeton: Princeton University Press.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.