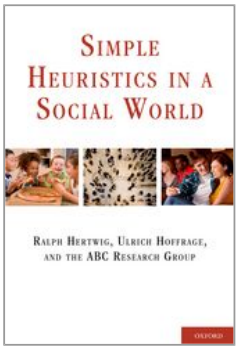
University Press Scholarship Online
Oxford Scholarship Online



Simple Heuristics in a Social World
Ralph Hertwig, Ulrich Hoffrage, and ABC Research Group

Print publication date: 2012
Print ISBN-13: 9780195388435
Published to Oxford Scholarship Online: January 2013
DOI: 10.1093/acprof:oso/9780195388435.001.0001

Simple Heuristics in a Social Game

Ralph Hertwig
Urs Fischbacher
Adrian Bruhin

DOI:10.1093/acprof:oso/9780195388435.003.0002

Abstract and Keywords

The dominant behavior observed in social games such as the ultimatum game, the dictator game, and public good games violates the classical assumption in economics of purely selfish preferences. To account for this behavior, economists have proposed social preference models, which introduce nonselfish motives as additional arguments and parameters in the utility function. Like classical utility models, social preference models focus on behavior at the expense of describing underlying cognitive processes, contenting themselves with being “as-if” models. This approach unnecessarily limits the models' psychological realism and forgoes the empirical benefits of describing the processes that produce behavioral outcomes. As an alternative, the chapter proposes fast and frugal classification trees. Designed to describe deliberations and decisions in the mini-ultimatum game, the trees spell out the possible cognitive processes of four distinct types of

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

respondents. The chapter derives response-time predictions from these trees as well as from a process interpretation of an influential social preference model, the Fehr and Schmidt model of inequity aversion, and test the predictions empirically. The observed response times suggest that a substantial proportion of respondents in the mini-ultimatum game take several distinct social considerations into account and process them sequentially, consistent with the proposed classification trees. The chapter discusses the implications of these findings for theories of economic behavior.

Keywords: economic games, ultimatum game, social preference models, fast and frugal classification trees, heterogeneity, response time, as-if models, psychological realism, kindness, social status

> Reasonable persons … desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept.
>
> *John Rawls (1993)*
>
> In the little world in which children have their existence … there is nothing so finely perceived and so finely felt, as injustice.
>
> *Charles Dickens (1860–1861/2003)*
>
> First, my pledge: More than ninety-nine percent of my wealth will go to philanthropy during my lifetime or at death. Measured by dollars, this commitment is large. In a comparative sense, though, many individuals give more to others every day. Millions of people who regularly contribute to churches, schools, and other organizations thereby relinquish the use of funds that would otherwise benefit their own families. The dollars these people drop into a collection plate or give to United Way mean forgone movies, dinners out, or other personal pleasures. In contrast, my family and I will give up nothing we need or want by fulfilling this ninety-nine percent pledge.
>
> *(Source: http://givingpledge.org)*

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

With this pledge, the legendary investor Warren E. Buffett—according to the *Forbes* list of the world's billionaires, the world's third-richest man in 2010 (Kroll & Miller, 2010)—announced that he would give away nearly all of his fortune and went on to challenge other billionaires to do the same. He also observed that, although his philanthropy is staggeringly generous in absolute terms, charitable giving by people less fortunate than he deserves more recognition because their donations to others involve real material sacrifices.

(p.40) Not only Warren Buffett but human beings in general display a wide array of social behaviors that would baffle anyone who assumes, as “most economists still routinely” do (Fehr & Schmidt, 2006, p. 617), that “material self-interest is the *sole* motivation of *all* people” (p. 617). According to Sen (2009, p. 184), this assumption “has come to dominate much of mainstream economic theory,” notwithstanding the fact that prominent and influential members of the economic hall of fame—including Adam Smith (1759/2010) and Nobel Prize laureates Gary Becker (1974), Kenneth Arrow (1981), Paul Samuelson (1993) and Amartya Sen (2009)—emphasized that people *do* care for the well-being of others.

One need not refer to exceptional philanthropists and Good Samaritans (Waldron, 2003) to question the veracity of the “assumption of the completely egoistic human being” (Sen, 2009, p. 184). All of us exhibit behaviors that conflict with the equation of human motivation and material self-interest. We tip in restaurants even when we are traveling in a foreign country and do not expect to be served by the same waiters ever again. We pay for fresh produce displayed by the roadside even when no one is watching (Dawes & Thaler, 1988). In short, we cooperate with strangers whom we will never meet again and in situations in which benefits from gains in reputation will be minuscule or absent. At the same time, we humans are also known to be capable of being nasty to others and willing to take revenge. Paradoxically, when spiteful behaviors are displayed in situations affording no prospect of material gain, they can also qualify as manifestations of unselfishness. In fact, our proclivity for punishing unfair and uncooperative behavior and our

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

willingness to do so without prospect of material advantage epitomize unselfish behavior because it probably helps enforce and uphold social order. It has even been proposed that altruistic punishment might be the glue that holds societies together and helps sustain cooperation (e.g., Boehm, 1993; Fehr & Gächter, 2002).

One possible explanation of unselfish behaviors is that, contrary to the portrayal of people as being exclusively driven by the goal of maximizing their self-interest, we are also guided by unselfish motives. Yet classic economic theory has a straightforward alternative explanation that retains the selfish foundations of human motivation. If there is a chance that one will encounter the same person again, and if nice people can expect to be treated nicely by other nice people, then unselfish behavior is a selfish person's strategic response to an expected future interaction. By this argument, people are prepared to punish others for treating them unfairly in order to secure fairer treatment in future interactions.

By merely observing naturally occurring interactions, it would be difficult to distinguish between the view that human motivation encompasses truly unselfish motives and the view that unselfish (p.41) behavior is part of the strategic repertoire of an agent maximizing his self-interest. However, by investigating highly scripted interactions in the laboratory, we can eliminate the prospect of future benefits and thereby test the two views. In what follows, we investigate unselfish behavior in a variant of what is perhaps the most important *drosophila* in the laboratories of experimental economics, the ultimatum game (Güth, Schmittberger, & Schwarze, 1982).

The Ultimatum Game: Peeking Into Human Nature

In its simplest form, the ultimatum game involves two people who play a single round in which one person, the *proposer*, suggests how to split a fixed monetary pie. This split represents a take-it-or-leave-it offer (an ultimatum) that the other person, the *responder*, must accept or reject. The interaction between the parties is anonymous. If the offer is accepted, the proposed division will be implemented. If the

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

offer is rejected, responder and proposer will go away empty-handed.

A rational and self-interested responder accepts any positive payoff, no matter how small. Aware of this, a rational and self-interested proposer will offer the smallest amount possible. The equilibrium offer (i.e., the offer for which no player has anything to gain by doing something differently) thus allocates the smallest positive payoff to the responder and the balance of that amount to the proposer. Predicting that this equilibrium will be reached, however, requires assuming that the responder is rational and acts selfishly *and* that the proposer believes the responder is rational and selfish, which is tantamount to assuming that she will accept any positive offer. A high offer thus does not contradict the notion of a selfish preference on the part of the proposer, because a proposer may reasonably believe an anonymous responder to be lacking in rationality and selfishness. In contrast, rejection of any positive offer is considered to be irrational on the part of the responder with respect to the norm of maximization of self-interest, because such a responder simply misses an opportunity to reap gains that come for free.

More than thirty years of research on the ultimatum game has hammered home one message: Low offers are frequently rejected (see Camerer, 2003; Güth & Tietz, 1990). In other words, this simple social game consistently triggers behavior that is at odds with the assumption that responders are rational and maximize self-interest. How can the numerous demonstrations that many people behave unselfishly or in an "other-regarding" way (two terms that we use interchangeably) be explained?

One leading response in economics has been to modify the neoclassical utility function, which depends only on self-regarding payoffs, by assuming that people also have other-regarding motives. (p.42) Psychological realism is thus introduced via additional terms in the utility function (in addition to the standard term reflecting the utility of the material payoff), while keeping the classical utility theory intact. This approach represents a "repair" program—in Selten's words (2001)—in that it strives to preserve the utility

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

framework by squeezing more into the utility function. Repair models have been proposed to explain choices in games against nature that conflict with maximization of expected utility (see Brandstätter, Gigerenzer, & Hertwig, 2006) as well as behaviors in social games that conflict with maximization of self-interest. Models for the latter class of behaviors are often referred to as *social preference models* (which Fehr & Schmidt, 2006, termed “theories of other-regarding preferences”). Before we illustrate these models in more detail, let us first introduce the social world we will focus on hereafter.

Our scripted social world represents a variant of the ultimatum game known as the *mini-ultimatum game*. Mini-ultimatum games are sequential two-player games. Player P (the proposer) first gets to choose between two possible divisions for both players. Player R (the responder) then chooses either to accept or to reject the division selected by the proposer. If the division is accepted, it will be implemented. If it is rejected, both players will receive nothing.

To use the established terminology in economics, the responder is presented with four payoffs: (P_o, R_o vs. P_f, R_f), with P_o and R_o denoting the division that the proposer selected and thus *offered* to the responder, and P_f, and R_f denoting the payoffs of the *forgone* division. If the responder accepts the proposed division, then P_o represents the proposer's income and R_o the responder's income. If the responder rejects the proposed allocation, both parties go away empty-handed.

The mini-ultimatum game is an ultimatum game in that the responder faces a take-it-or-leave-it offer. Yet it is also both more restrictive and more general than the ultimatum game. It is more restrictive in that it limits the number of possible divisions that the proposer can suggest to two. It is more general in that the size of the monetary pie can vary across proposed divisions, as in the following two: 3, 5 versus 2, 8 (the total pie is thus 8 and 10, respectively).

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Social Preference Models

Social preference models differ with respect to the unselfish motives that they assume. One important motive is *inequity aversion* (e.g., Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). According to Fehr and Schmidt's model, for instance, people dislike inequity. In addition to the absolute size of their own payoff, they care about its relation to that of the other player. When their own payoff exceeds that (p.43) of the other player, they are willing to lessen the inequity by, for instance, transferring money to the other player. Once the other player's payoff exceeds her own, however, the player—say, the responder in the ultimatum game—feels envious and rejects low offers. On this view, the responder's decision to accept or reject hinges solely on the (in)equity of the proposed division (i.e., P_o vs. R_o), with the forgone division playing no role in the deliberations. If the inequity between the proposer's and the responder's payoffs is large relative to the responder's payoff, the division will be rejected; otherwise it will be accepted.

Experimental findings, however, suggest that at least some people do take the forgone division into account (Brandts & Sola, 2001; Falk, Fehr, & Fischbacher, 2003). *Intention-based reciprocity* models stress the importance of this "counterfactual world" (Dufwenberg & Kirchsteiger, 2004; Rabin, 1993) by assuming that, to varying extents, players care about their counterpart's intentions and want to pay back in kind. That is, if a responder senses that the proposer wants to treat him kindly, he will want to be kind in return. If he perceives the proposer as harboring hostile intentions, he will want to teach her a lesson. In this view, the key determinant of a responder's behavior is his interpretation of the proposer's behavior. Different models define the proposer's kindness toward the responder in different ways. Rabin's definition is the most influential: In the standard ultimatum game, kindness is measured by the difference between the actual payoff that the proposer offers the responder and the "fair" payoff (defined as the middle of the range of the responder's possible payoffs). In the context of the mini-ultimatum game, the model predicts punishment (rejection) when the

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

responder's payoff—given the proposed alternative—is smaller than in the forgone alternative, as in the following game: 8, 2_o versus 5, 5 (with *o* representing the offered division).

Although attempts to explain other-regarding behavior within the classic utility framework are not limited to inequity aversion and intention-based reciprocity models (for an extensive review, see Fehr & Schmidt, 2006), they share with all social preference models the assumption that motivations vary from person to person. Consequently, the assumption that some people are concerned with fairness and others with kindness and reciprocity does not mean that all people share these concerns equally. Social preference models accommodate such heterogeneity in terms of adjustable parameters representing the strength of the various other-regarding concerns. For instance, Fehr and Schmidt's (1999) theory of inequity aversion assumes a utility function that includes two adjustable inequity-aversion parameters, with α representing disadvantageous inequity aversion (i.e., how much a person dislikes having a lower payoff than someone else) and β representing advantageous inequity aversion (p.44) (i.e., how much a person dislikes having a higher payoff than someone else). These parameters can take on different values for different individuals, reflecting the weight that a particular person bestows on these social concerns; they may, however, just as well turn out to be zero, meaning that the person is focused exclusively on maximizing her own payoff.

To summarize, some behaviors observed in economic experiments consistently violate the classical assumption of purely selfish preferences. To account for this behavior, economists have proposed social preference models that introduce unselfish motives as additional arguments and parameters in the utility function. Heterogeneity among people's unselfish concerns is accommodated via adjustable parameters whose values reflect how strongly the forces that the parameters represent influence a person's behavior.

Despite their success in explaining overt behavior, social preference models do not answer Rubinstein's recent call (2003) “to open the black box of decision making” (p. 1215). Like classical utility models, they focus on behavior at the

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

expense of describing underlying cognitive processes, thus contenting themselves with being *as-if* models (Berg & Gigerenzer, 2010). This approach unnecessarily limits the models' psychological realism and the empirical content that can be gained by describing the processes that produce behavioral outcomes. According to Rubinstein, to move beyond *as-if* models, one would need to "come up with some completely new and fresh modeling devices" (p. 1215). One fresh approach, or so we believe, is to understand not only *what* responders in the mini-ultimatum game do, but *how* they do it in terms of boundedly rational heuristics.

Models of Bounded Rationality: Fast and Frugal Classification Trees

With an aim of explaining human behavior in terms of boundedly rational strategies such as heuristics, Gigerenzer and Selten's book *Bounded Rationality: The Adaptive Toolbox* (2001) promoted "bounded rationality as the key to understanding how actual people make decisions without utilities and probabilities" (p. i). In their view, quantitative probabilities, utilities, and even utilities enriched by psychological concepts such as kindness or inequity aversion fail to capture the essence of the processes that actually unfold in the human mind, whereas heuristic processes such as recognition, limited search, stopping rules, and one-reason decision making do. Boundedly rational heuristics play a crucial role in *fast and frugal classification trees*, which represent a genuinely new approach to modeling decisions about classification (Martignon, Katsikopoulos, & Woike, 2008; 2012)—one of the most important cognitive tasks that we face.

(p.45) A fast and frugal tree is a classification device that could yield a binary classification at each level of the tree. For illustration, consider the following tree designed by Green and Mehr (1997) to model physicians' decisions about whether to admit patients suffering chest pain to the coronary care unit (CCU) in U.S. hospitals. A man with serious chest pain is rushed to the hospital. Suspecting acute ischemic heart disease, the doctors need to decide quickly whether he should be admitted to the CCU or assigned to a regular nursing bed.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

This decision is not trivial, as the CCU is a highly strained resource due to the practice of defensive medicine (chapter 3). To model how doctors make the decision, Green and Mehr developed a fast and frugal classification tree that asks a maximum of three questions.

As illustrated in Figure 2-1, the first question is whether the patient's electrocardiogram (ECG) reveals a specific anomaly in the ST segment, a component of the ECG tracing of the cardiac cycle. If the answer is yes, he is at once admitted to the CCU without consulting other diagnostic information. If the answer is no, the next question is posed: Is the patient's chief complaint chest pain? If the answer is no, he is assigned to a regular nursing bed. If the answer is yes, a final, composite question is used to classify the remaining patients. Forgoing calculation of probabilities, the tree processes a maximum of three criteria one at a time. Each criterion enables the physician

(p.46) either to render a final judgment or to turn to the next criterion. More generally, a classification tree has *M* decision criteria. Each criterion is associated with one end node, except for the last criterion, which has two end nodes (i.e., each tree has a total of M + 1 end nodes).

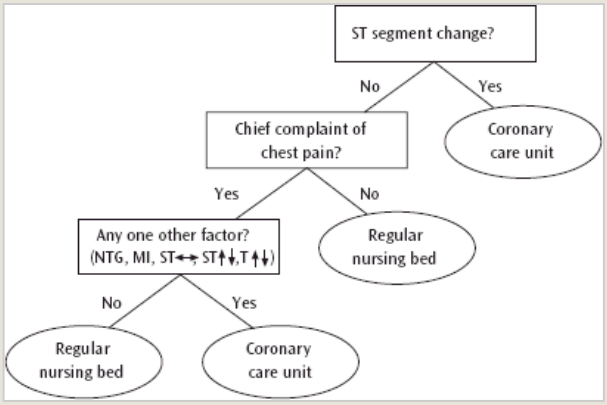


Figure 2-1: A fast and frugal classification tree for deciding whether a patient with chest pain will be assigned to the coronary care unit or a regular nursing bed. (Based on “What alters physicians’ decisions to admit to the coronary care unit?” by L. Green & D. R, Mehr, *Journal of Family Practice*, *45* (1997): 219–226.)

Classification Trees in the Mini-Ultimatum Game

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

The responder's decision in the mini-ultimatum game can be thought of as a classification. Specifically, the responder *classifies* the division in question as one that he deems "acceptable" or "unacceptable" and responds accordingly. Before we can design specific trees as process and outcome models of responders' decisions in the mini-ultimatum game, however, we need to answer two questions. First, what criteria could people use to decide whether to accept or reject a division? Second, which criteria matter to whom?

Which Criteria May Be Consulted in the Mini-Ultimatum Game?

Each mini-ultimatum game presents the responder and the proposer with four payoffs, two of which represent the proposed division and two the forgone division. Using these four payoffs, we focus on three criteria that figure prominently in social preference models, and put forward a fourth, novel, one.

Positive payoff. According to this criterion, a division will be accepted if the responder's payoff (R_o) is positive (i.e., $\rangle$ 0), no matter how small the offered amount is. This is the sole criterion of a perfectly rational and selfish player, who considers anything better than nothing.

Status. According to this criterion, a division will be accepted if the responder's payoff is at least as large as the proposer's (i.e., $R_o \, ^3 \, P_o$). This criterion embodies social comparisons, which in social psychology have long been recognized as important (see e.g., Festinger, 1954; Mussweiler, 2003) and form the core of *inequity aversion* theories (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999).

Kindness. According to this criterion, a division will be accepted if the responder's payoff is at least as large as the payoff she would have received had the proposer selected the other division (i.e., $R_o \, ^3 \, R_f$). Put differently, the responder checks whether the proposer has chosen the "kinder" option. This criterion is the core of several

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

reciprocity models (Dufwenberg & Kirchsteiger, 2004; Rabin, 1993).

Mirror. According to this newly proposed criterion, a division will be accepted if the responder would have made (p.47) the same proposal had he been the proposer and faced the same choice. The mirror criterion embodies the responder's attempt to walk a mile in the shoes of the proposer, simulating the dilemma that the choice may present to the proposer.

For illustration of the mirror criterion, consider a proposer who must choose between two highly unequal divisions, neither appearing to be fair: 8, 2 and 2, 8. Knowing the proposer's dilemma, a responder can gauge her own (hypothetical) proposal by taking an honest look in the mirror (hence the “mirror” criterion), using her own inferred behavior to decide whether a division is acceptable. Adopting one's inferred behavior as a benchmark can be seen as norm-abiding behavior in an environment with heterogeneous norms (see, e.g., López-Pérez, 2008), in which people's notions of fairness are likely to differ: Even in such environments, people who apply the mirror criterion are internally consistent in that they comply with their own standards and—by rejecting some offers—punish those who propose a division that they would not have chosen.

It should be acknowledged that, like social preference models, the mirror criterion is a black box—one we leave closed here. In addition, unlike those of the three criteria outlined above, its behavioral predictions cannot simply be derived from the objective payoffs. They can only be empirically derived, separately for each individual player. We employed two methods for generating player-specific predictions. One uses people's actual (and thus observable) behavior as proposers in the same mini-ultimatum games, the other their actual behavior in equivalent “dictator” games, as a proxy for the mirror criterion. We return to both these methods and to the topic of dictator games shortly.

Figure 2-2 depicts the four criteria. Of course, the responder could in theory evaluate the four payoffs in any mini-

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

ultimatum game by other criteria or other combinations of the payoffs, but many of those would not be psychologically meaningful. For instance, the two diagonal conjunctions (i.e., P_o vs. R_f and R_o vs. P_f) make little sense because they do not compare payoffs within one player or within one division. Similarly, considering just one payoff appears meaningful only in the case of R_o.

Who Will Consult Which of the Criteria?

There are at least two approaches to answering this question, one theoretical and the other empirical. We took the empirical approach. Specifically, we conducted an experiment including a set of 12 mini-ultimatum games (see Box 2-1). In each game, we recorded the responders' decisions and, importantly, the time they needed to decide. (p.48)

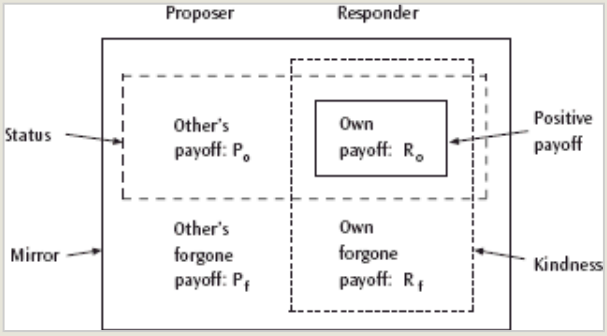


Figure 2-2: A mini-ultimatum game provides four payoffs: the payoffs that the proposer (P_o), and responder (R_o) would receive if the offered division were accepted, and the payoffs that the proposer (P_f) and the responder forgo (R_f). Four different subsets of the four payoffs constitute four criteria, embodying selfish (positive payoff) and social concerns (status, kindness, and mirror) to determine acceptance or rejection of an offer (see text).

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Box 2-1: Experimental Task

The experiment included 12 mini-ultimatum games in which three properties of the divisions were varied: inequality, kindness, and efficiency (see Table 2-1). We elicited decisions using the strategy method. According to this method, the responder makes a decision for each of the divisions constituting one game (e.g., 5, 5 vs. 8, 2). For instance, a responder is asked to assume that (5, 5) has been offered and (8, 2) forgone. Later, she is asked to assume that (8, 2) has been offered and (5, 5) forgone. Thus, there were two possible divisions per game, and participants made acceptance or rejection decisions for a total of 24 offers across 12 games. The order of offers was randomized, with the constraint that the two offers from a given game were not presented in succession.

Seventy students from the Swiss Federal Institute of Technology Zurich and the University of Zurich participated. Sessions lasted between 60 and 75 minutes. The experiment was programmed and conducted using z-Tree software (Fischbacher, 2007). Each participant first made 24 decisions assuming the role of a responder. Next, he made decisions for the same 12 games (Table 2-1), first in the role of the proposer and then in the role of a "dictator" (see text). The sequence of the three blocks—responder, proposer, and dictator decisions—was always the same; within each block, however, the sequence of games was randomized.

For each participant, we randomly selected one game in which she was a proposer and one in which she was a responder, respectively. For both games, the participant was randomly teamed up with another player, and both were paid according to their respective decisions. Finally, we paid each participant for one of his dictator decisions. Participants were fully informed about how their ultimate payment was determined. On average, they earned 39.65

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Swiss francs (about $34), including a participation fee of 10 Swiss francs.

(p.49)

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Table 2-1: The Mini-Ultimatum Games Studied in the Experiment (See Box 2-1)

Number of game	Proposer	Responder	Proposer	Responder	Number of game
	Offered distributions (for games 1–12)		Forgone distributions (for games 1–12)		
1	8	8	2	5	13
2	4	8	2	5	14
3	4	5	1	6	15
4	8	5	8	2	16
5	5	5	8	2	17
6	4	6	8	2	18
7	2	5	1	8	19
8	6	4	8	2	20
9	5	3	8	2	21
10	1	8	2	5	22
11	2	8	8	2	23
12	8	4	2	2	24

Note. Games were selected to represent different combinations of efficiency ("efficiency" refers to the combined payoff of the two players), inequality, and kindness. Each game was presented in two variants, one in which the pair of payoffs in the left columns (e.g., Game 1: 8, 8) represented the offered divisions, and the payoffs in the right columns represented the forgone divisions (Game

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

1: 2, 5), and vice versa (Games 13–24). Participants saw these amounts, multiplied by a factor of 100 (that is, 800, 800 vs. 200, 500 in Game 1).

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Let us focus first on the decisions. The decisions obtained from the responders were used (a) to examine whether the four criteria indeed entered people's deliberations and (b) to find out who consulted which criteria.

First, when both the status criterion and the kindness criterion supported acceptance of the proposer's offer, only 3% of divisions were rejected. In contrast, when both criteria suggested rejection, the rejection rate climbed to 30%. Finally, when only the status criterion or only the kindness criterion suggested rejection, 6% and 5% rejections were observed, respectively. Taken together, these results show that a division was much more likely to be accepted when the responder received at least as much as the proposer or at least as much as he would have in the forgone division than in a division that failed both the status and the kindness tests.

Second, contrary to the mirror criterion, a person's decisions as a proposer in the mini-ultimatum game were not predictive of her decisions when she played the same game as a responder. This suggests that proposer behavior in the mini-ultimatum game is more driven by the fear of rejection than by sentiments about the fairness or equity of an allocation. Fortuitously, however, a person's decisions in the dictator game proved to be predictive of his behavior as a responder. In a dictator game, players determine which of the two (p.50) divisions they would propose assuming that the pie is split accordingly, without any threat of a responder's retaliating. Therefore, a dictator need make no assumptions regarding the other player's preferences, and so, presumably, only fairness considerations motivate the proposal of equitable divisions in dictator games (e.g., Forsythe, Horowitz, Savin, & Sefton, 1994).

Finally, using a regression analysis, we confirmed that the three criteria that the rational and selfish player should ignore —status, kindness and mirror—in fact matter (for detailed results and statistics, see Fischbacher, Hertwig, & Bruhin, 2012). Based on these results, we conclude that the three social criteria we have identified mattered to our responders. But which of them mattered to whom? To answer this question, we entered responders’ rejection rates in our

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

experiment (Box 2-1) into a finite-mixture-model analysis (McLachlan & Peel, 2000; see also Fischbacher et al. for more details) to define a finite set of models and, if possible, to assign each participant to a model. More formally, each individual was assigned a probability distribution over the models on the basis of which (if sufficiently discriminative) a participant was classified into distinct models of responders. In what follows, we describe the models of responder behavior identified in the finite-mixture-model analysis.

Figure 2-3 depicts the four responder models in terms of their average rejection rates across the mini-ultimatum games employed (Table 2-1). The rejection rates are highly variable across models. Consider, for instance, Model 1: A substantial proportion of responders—30 out of 70 (43%)—behaved consistently selfishly in that they *never* rejected a single proposed positive allocation, no matter how small.

(p.51)

Responders subsumed under Model 2 ($N = 6$), in contrast, accepted all proposed divisions, except when the unfavorable inequality between payoffs (R_o vs. P_o) became highly pronounced.

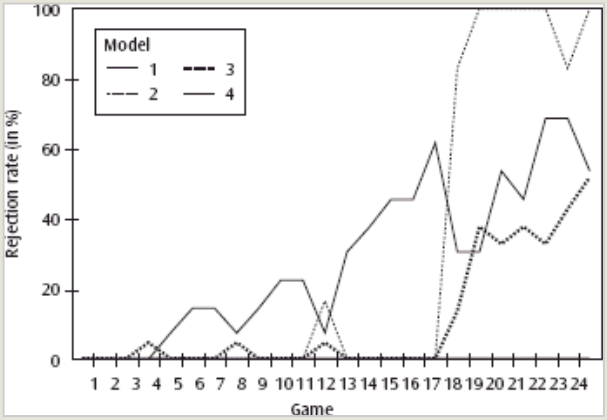


Figure 2-3: Rejection rates across the 24 proposed divisions (Table 2-1), separately for models of responders.

When the inequality was too large, as in the case of the proposed division 8, 2, the rejection rate was nearly 100%, irrespective of the forgone division. As postulated in *inequity aversion* theories, Model 2 responders' overriding consideration appeared to be their status relative to that of the proposer. Similarly, Model 3 responders ($N = 21$) accepted all allocations in which the payoff inequality was not too pronounced. Rather than categorically rejecting ostensibly unequal offers, however, they made their rejections contingent on

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

the forgone division. That is, as suggested in intention-based reciprocity theories (Dufwenberg & Kirchsteiger, 2004; Rabin, 1993), they appeared to consider the proposers' intentions, tending to reject unkind offers in which their actual payoff was worse than the forgone one (R_o vs. R_f). Finally, Model 4 responders ($N = 13$) rejected the widest range of gambles, although the rates of rejection did not approach 100%. No apparent regularity can be gleaned from their behavior. These responders might have been using the individual-specific mirror criterion, a possibility to which we return shortly.

Based on these four responder models, we could examine which of the criteria are likely to have shaped the behavior captured by each model. To this end, we classified the mini-ultimatum games into eight distinct profiles defined by the predictions of the criteria, separately for each model. The letters "A" and "R" represent the predictions of a given criterion (acceptance vs. rejection). Table 2-2 reports the profiles and associated rejection rates. The right-most column reports the rejection rates averaged across responders assigned to the respective model; the criteria profiles are ordered according to this quantity. We were now able to analyze which criterion or set of criteria best maps onto the observed rejection rates. For instance, Model 1 decisions were perfectly aligned with the predictions of the positive-payoff criterion.

Model 2 responder behavior is more differentiated. The first four classes of games are consistent with both the predictions of the positive-payoff criterion and the status criterion (Table 2-2). The last four classes of games, however, conflict with the positive payoff criterion. Although the status criterion likewise cannot predict them, it discriminates between the first four classes (acceptance) and the last four classes (rejection). Therefore, combined with an additional criterion, it may have determined these responders' behavior. Indeed, the kindness criterion maps reasonably well onto the observed rejection rates of the last four classes of games. A similar albeit less clear pattern emerges for Model 3 responders. Finally, Model 4 responders show the most heterogeneous rejection behavior. Admittedly, their medium-size rejection rates (ranging between 16% and 58%) are hard (p.52)

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Table 2-2: Rejection Rate as a Function of Four Models of Responders (See Text) and Four Criteria

Model of responder	Criteria				Number of decisions	Rejection rate%
	Positive payoff ($R_o > 0$)	Status ($R_o \geq P_o$)	Kindness ($R_o \geq R_f$)	Mirror		
1	A	R	R	A	92	0
1	A	A	A	A	96	0
1	A	R	A	A	113	0
1	A	A	R	A	59	0
1	A	A	A	R	144	0
1	A	R	A	R	37	0
1	A	A	R	R	91	0
1	A	R	R	R	88	0
2	A	A	R	A	15	0
2	A	A	A	A	28	0
2	A	A	R	R	15	0
2	A	A	A	R	20	0
2	A	R	A	A	24	13
2	A	R	A	R	6	50

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Model of responder	Criteria				Number of decisions	Rejection rate%
	Positive payoff ($R_o > 0$)	Status ($R_o \geq P_o$)	Kindness ($R_o \geq R_f$)	Mirror		
2	A	R	R	R	31	97
2	A	R	R	A	5	100
3	A	A	A	A	81	0
3	A	A	R	A	41	0
3	A	A	A	R	87	0
3	A	A	R	R	64	0
3	A	R	A	R	22	0
3	A	R	A	A	83	7
3	A	R	R	A	47	30
3	A	R	R	R	79	46
4	A	A	A	A	57	4
4	A	R	A	A	49	6
4	A	A	R	A	30	17
4	A	R	A	R	16	31
4	A	R	R	A	20	40

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Model of responder	Criteria				Number of decisions	Rejection rate%
	Positive payoff ($R_o > 0$)	Status ($R_o \geq P_o$)	Kindness ($R_o \geq R_f$)	Mirror		
4	A	A	A	R	47	40
4	A	A	R	R	35	40
4	A	R	R	R	58	59

Note. “A” denotes that the respective criterion predicts that the offered payoff division will be accepted; “R” denotes that the respective criterion predicts rejection of the division

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

to account for because all criteria are binary predictors. Nevertheless the qualitative changes (i.e., lower vs. higher rejection rates) are best fitted by the mirror criterion. When the mirror criterion predicts rejection, the average rejection rate equals 43%, compared with 17% when it predicts acceptance. No doubt there is some additional heterogeneity (or random variation) that neither the mirror criterion nor combinations of other criteria can capture. Of the four criteria, however, the (p.53) mirror criterion best tracks the empirical variations in rejection rates among Model 4 responders.

Four Classification Trees

Based on the classification of responders into four models and the results in Table 2-2, we designed four classification trees.

Selfish Tree

This tree maps onto the behavior of Model 1 responders. It embodies one-reason decision making (Gigerenzer, Todd, & the ABC Research Group, 1999) in that only one criterion is consulted:

> Step 1. Look up your payoff (R_o).
> Step 2. Accept the allocation if R_o is larger than 0; otherwise reject.

This tree correctly predicts 100% of the 720 decisions made by Model 1 responders.

Priority Tree

This tree maps onto the behavior of Model 2 responders. It consists of three criteria: the positive payoff criterion, the status criterion, and the kindness criterion, with status taking priority over kindness:

> Step 1. Look up your payoff (R_o).
> Step 2. If your payoff is larger than 0, go to Step 3; otherwise reject.
> Step 3. Look up the proposer's payoff (P_o).
> Step 4. Accept if your payoff is at least as large as the proposer's payoff; otherwise go to Step 5.
> Step 5. Look up your forgone payoff (R_f).

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Step 6. Accept if your actual payoff is at least as large as the forgone payoff; otherwise reject.

This tree looks up a minimum of one criterion (if $R_o = 0$) and a maximum of three criteria. It correctly predicts 95% of the 144 decisions made by Model 2 responders.

Priority+1 Tree

This tree maps onto the behavior of Model 3 responders. It is identical to the priority tree, except that that Step 6 is replaced by the following sequence:

Step 6. Accept if your actual payoff is at least as large as the forgone payoff; otherwise go to Step 7.
Step 7. Look up both divisions (i.e., P_o and R_o vs. P_f and R_f).
Step 8. Accept if you would have selected the same allocation had you been the proposer; otherwise reject.

(p.54) The priority+1 tree correctly predicts 88% of the 504 decisions made by Model 3 responders. Admittedly, adding the mirror criterion captures some, but by no means all, of the variance: The rate of acceptance is 70% when the mirror criterion predicts acceptance, compared with 54% when it predicts rejection.

Mirror Tree

This tree maps onto the behavior of Model 4 responders, who appear to take the proposer's perspective, considering the temptations and trepidation that the other player may experience when confronted with a choice between two specific divisions:

Step 1. Look up your payoff (R_o).
Step 2. If your payoff is larger than 0, go to Step 3; otherwise reject.
Step 3. Look up both divisions (i.e., P_o and R_o vs. P_f and R_f).
Step 4. Accept if you would also have selected the proposed division had you been the proposer; otherwise reject.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

The mirror tree correctly predicts 67% of the 312 decisions made by Model 4 responders. This performance is obviously much lower than that obtained for the other trees. Without the mirror criterion, however, we would not have been able to describe these responses at all.

To summarize, all four trees are simple and transparent models of the possible deliberations different types of responders may engage in to arrive at a decision. The selfish tree, which includes only one criterion (the positive-payoff criterion), does not take into account the forgone payoff or the payoffs of the other player. The social trees are more differentiated. Specifically, the priority and the priority+1 trees include three and four binary questions, respectively, that are sequentially ordered. Both trees implement one-reason classification (Martignon et al., 2008; 2012), making at least one classification for each criterion (either reject or accept). Finally, the mirror tree includes two criteria (positive payoff and mirror), of which the mirror criterion is the cognitively most demanding of all criteria in requiring the willingness and ability to adopt another person's perspective.

Two disclaimers are in order. First, note that, unlike the social preference models, the fast and frugal trees use binary criteria rather than thresholds or explicit tradeoffs among criteria (see also Brandstätter et al., 2006), although it seems plausible that small losses in status or kindness are less likely than large ones to cause somebody to reject a proposal. Second, we designed the trees by *fitting* them to responders' choices post hoc, after having classified responders into one of four models using the finite-mixture-model analysis. Moreover, the acceptance rates at the terminal nodes of the (p.55) trees do not always equal, or even approximate, either zero or one hundred percent. That is, the trees do not fit perfectly, and people's deliberations may have been influenced by criteria not considered here and perhaps also by random variation.

Response Times: A Window Onto Psychological Processes

We have described two frameworks to account for behavior in mini-ultimatum games that represent fundamentally different

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

modeling strategies. The classification trees explicitly spell out the processes, whereas, as black boxes, the social preference models refrain from invoking specific and empirically supported psychological processes. Notwithstanding social preference models' exclusive concern with observable behavior, we attempted to interpret them in terms of a process by taking advantage of response times.

Why response times? In psychology, the analysis of response times has often been used to distinguish among models of decision strategies that are difficult or impossible to tease apart on the level of choices. To this end, experimenters have devised various process-tracing methods (for a review, see Reisen, Hoffrage, & Mast, 2008), such as the tracking of eye movements (e.g., Rayner, 1998), think-aloud protocols (e.g., Ericsson & Simon, 1993) and the Mouselab method (e.g., Payne, Bettman, & Johnson, 1993; for a recent application in economics, see Gabaix, Laibson, Moloche, & Weinberg, 2006). Unlike other process-tracing methods that more explicitly monitor information search or the allocation of attention (e.g., Mouselab), the analysis of response times offers a window onto the underlying cognitive process without the risk of altering the behavior that it is intended to measure (Bergert & Nosofsky, 2007).

Classification Trees and Response Times

Let us first consider the response-time predictions that can be derived from the four classification trees. Because the trees rest on sequential processing of criteria and yield a decision as soon as a criterion is met, one can predict that the more criteria people examine, the more time they need to make a choice. Furthermore, it is plausible to assume that the more complex a criterion is, the longer it takes to examine it. One simple measure of complexity is the number of payoffs (i.e., P_o, R_o, P_f, and R_f) that must be looked up and compared. By this measure, the least complex criterion is the positive-payoff criterion (requiring R_o to be looked up and compared to 0), and the most complex, the mirror criterion (requiring all four payoffs to be looked up and compared). From this notion of

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

complexity and the (p.56) nature of sequential processing, three response-time predictions can be derived:

> *Prediction 1. Making a decision using the selfish tree will require less time than using any of the other trees, regardless of whether the decision is to accept or reject.*

This prediction follows from the fact that the selfish tree examines only one criterion (positive payoff), which requires retrieval of a single piece of information (i.e., R_o); in contrast, all other trees require the examination of more than one criterion.

> *Prediction 2. Making a decision using the mirror tree will require more time than using the priority tree.*

This prediction follows from the assumption that inferring one's own hypothetical behavior is time-consuming because it requires retrieving all four payoffs as well as determining one's hypothetical preferences. In contrast, the priority tree requires—provided that the positive payoff criterion has been met—looking up at most all four payoffs (i.e., checking the status and kindness criteria) and possibly as few as two payoffs (i.e., checking only the status criterion). In addition, examination of each criterion requires merely a simple arithmetic operation (e.g., is $R_o \stackrel{?}{>} P_o$?).

> *Prediction 3. The more criteria that are examined within the priority and priority+1 trees, the more time will be required to make a decision.*

This prediction follows from the sequential processing of the criteria. Specifically, the priority tree predicts shorter response times when the first criterion (status) suggests accepting the allocation, relative to when the second criterion (kindness) also needs to be consulted. In addition, the priority+1 tree predicts that those decisions that require consulting the mirror reason will take longer than those that require consulting only the status criterion or the kindness criterion, respectively.

Social Preference Models and Response Times

How would social preference models explain heterogeneity in response times across responders and across games, if it indeed occurs? We start with the simple and plausible assumption that response time increases with cognitive complexity (e.g., Bergert & Nosofsky, 2007; Payne et al.,

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

1993). From this follow two predictions. (p.57) First, if responders invoke a utility calculation to evaluate the proposed division, and if they differ merely in how they weigh their concerns (self-interest and other-regarding motives), then all people should have the same response time or at least the same response-time pattern. Second, if for some people only one motive matters (e.g., self-interest), their calculations should be simpler than for people who trade off their self-interest and other-regarding motives (e.g., inequity aversion). Consequently, the response times of the former should be shorter than those of the latter, consistent with Prediction 1 (see above).

Let us illustrate what these qualitative predictions mean using Fehr and Schmidt's (1999) model of inequity aversion as well as their utility function capturing disutility from advantageous and disadvantageous inequality:

$$U_i = \pi_i - \frac{\alpha_i}{N-1}\sum_j \max[\pi_j - \pi_i, 0] - \frac{\beta_i}{N-1}\sum_j \max[\pi_i - \pi_j, 0].$$

According to this equation, disutility due to inequity is subtracted from the utility stemming from the material payoff π_i. Disadvantageous inequality has a weight of α, and advantageous inequality has a weight of β; N denotes the number of players (two in our case). Assuming that model complexity is a function of how many concerns need to be traded off, one can predict that for responders whose α and/or β parameters equal zero, the calculations are simpler. Consequently, their response times should be faster than those with nonzero α and β parameters—a prediction that is identical to Prediction 1.

Social preference models can thus point to response-time predictions that are in principle consistent with those of the classification trees. However, some predictions do discriminate between the two approaches. The most obvious example is Prediction 3, which states that the response time of the very same responder changes as a function of the game (i.e., the number of criteria that must be retrieved). Because social preference models evaluate all potentially relevant arguments in a given person's utility function, they cannot predict the dependency of response times on games postulated by Prediction 3.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

A Test of the Response-Time Predictions

We next tested Predictions 1, 2, and 3 against the observed response times. Response times were entered into regressions with robust standard errors that take account of the fact that a given individual's decisions are not independent of one another (because response times were log normally distributed, we used their logarithms).

(p.58) Prediction 1 states that making a decision using the selfish tree will require the shortest amount of time. To test this prediction, we compared the response times of responders subsumed under Model 1 (i.e., “users” of the selfish tree) with those of responders subsumed under Models 2, 3, and 4. Consistent with Prediction 1, the average response time of selfish tree users was 3.5 seconds (Mdn = 2.4; SD = 3.9), compared with 5.2 seconds (Mdn = 4.3; SD = 3.9) for all other responders (regression of log(RT); $p < 0.001$).

There is one potentially confounding factor in this test of Prediction 1. The response times of selfish-tree users reflect only acceptances. The reason is that all the offered payoffs were larger than zero, and so (by definition) these responders never rejected a division. In contrast, the response times of users of the other trees include both acceptances and rejections. According to a finding by Knoch, Pascual-Leone, Meyer, Treyer, and Fehr (2006), rejecting a proposed ultimatum game allocation takes more time than accepting it. Consequently, the observed difference in response times between selfish-tree users and users of other trees could simply stem from the fact that the former group accepted every division, whereas the latter occasionally rejected a division. To control for this possibility, we reanalyzed the response times of unselfish-tree users (i.e., Models 2, 3, and 4 responders), *conditioning* on acceptance. Corroborating Prediction 1, their average response time still proved to be markedly longer than the 3.5 seconds observed for users of the selfish tree (M = 5.0; Mdn = 4.1; SD = 3.8; regression of log(RT): $p < 0.001$).

Finally, the response times of selfish-tree users were independent of whether the status and kindness criteria,

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

respectively, predicted acceptance, again confirming that the sole concern of selfish responders is their own payoff. In the dictator game, however, the response times of selfish responders (although, of course, not their choices) did depend on the decisions they made. When in the role of a dictator they did *not* select the division in question, their response times were longer than when they did select a division (not selected: $M = 3.9$, $Mdn = 2.7$, $SD = 4.2$; selected: $M = 3.1$, $Mdn = 2.2$, $SD = 3.4$ regression of log(RT): $p < 0.001$). One interpretation of this result is that even purely selfish players have other impulses that they appear to be able or willing to override consistently.

According to Prediction 2, making a decision using the mirror tree will take longer than using the priority tree. To test it, we compared the response times of responders subsumed under Model 2 (i.e., "users" of the priority tree) against those subsumed under Model 4 (i.e., "users" of the mirror tree). Consistent with Prediction 2, the average response time of priority-tree users was 4.2 seconds ($Mdn = 3.6$, $SD = 2.8$), relative to 6.2 seconds for mirror-tree users ($Mdn = 5.2$, $SD = 4.7$; regression of log(RT); $p = 0.04$). This difference remained when the analysis was limited to acceptance decisions (p.59) ($M = 3.8$, $Mdn = 2.6$, $SD = 2.5$ vs. $M = 5.9$, $Mdn = 5.1$, $SD = 4.7$; regression of log(RT): $p = 0.027$). Prediction 3 states that the more criteria that are examined within the priority and priority+1 trees, the more time will be required to make a decision. To test it, we examined the response times of responders

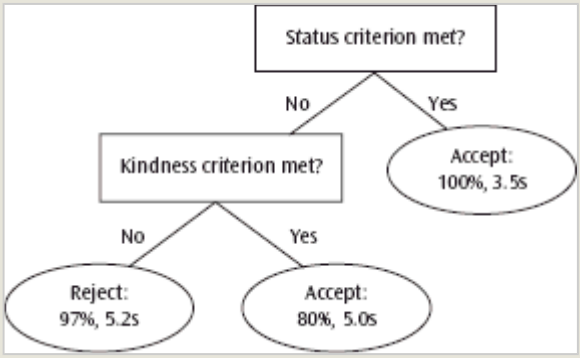


Figure 2-4: Correctly predicted decisions and average response times for the priority tree (Model 2 responders).

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

subsumed under Models 2 and 3 (i.e., users of the priority tree and the priority+1 tree), respectively. As Figure 2-4 shows, the average response time of Model 2 responders was indeed shorter in games where the status criterion suggested acceptance (M = 3.5, Mdn = 2.5, SD = 2.3) than when the kindness criterion also had to be consulted. When the kindness criterion suggested acceptance and rejection, the response times were M = 5.0 (Mdn = 3.8, SD = 3.6) and M = 5.2 (Mdn = 4.6, SD = 3.9), respectively; both differed significantly from the response time associated with the status criterion only (regression of log(RT): p = 0.037 and 0.003, respectively).

The part of Prediction 3 concerning the priority+1 tree was also largely confirmed. As Figure 2-5 shows, the average response time of Model 3 responders was slightly shorter in games where only the status criterion had to be consulted (M = 4.5, Mdn = 3.7, SD = 3.2) than when the kindness criterion had to be examined as well (M = 4.7, Mdn = 4.0, SD = 2.7). This difference, which is not statistically significant (regression of log(RT): p = 0.161), could reflect heterogeneity in the order in which the two criteria are processed. Clearly, however, consultation of the third and final criterion, the mirror criterion, requires more time: Relative to decisions in which the kindness criterion suggested acceptance, decisions made on the basis of the mirror criterion took longer (M = 6.1, Mdn = 5.0, SD = 4.3; regression of log(RT): p = 0.002).

Modeling of responder decisions in terms of classification trees also allows us to revisit a finding mentioned earlier. Knoch et al. (p.60)

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

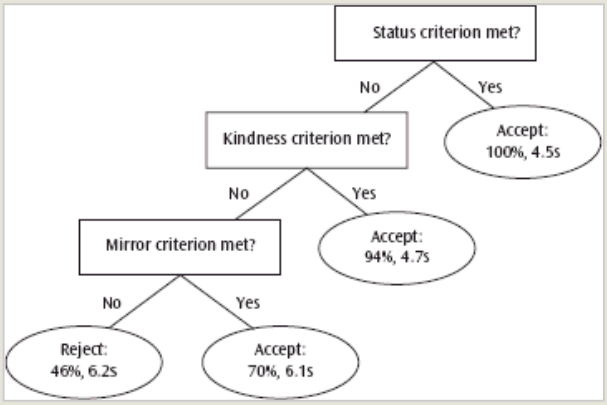


Figure 2-5: Correctly predicted decisions and average response times for the priority+1 tree (Model 3 responders).

(2006) observed that rejections in the ultimatum game take longer than acceptances. According to their explanation, when self-interest and social considerations such as fairness conflict with each other, overriding one's self-interested impulses takes time. Classification trees suggest an additional explanation; namely, that rejections take longer because they occur later in the sequence of decision steps. For instance, in the priority+1 tree, the only consideration that can yield rejections is the mirror criterion, which is consulted last. When not met, the preceding criteria lead, not to rejections, but to “passing the buck” to the next criterion. The only exception is the positive-payoff criterion, which can predict rejection if the offered payoff is zero; however, in our mini-ultimatum games a zero payoff was never offered (Table 2-1). Thus, it may not only be the conflict between self-interest and social considerations per se that makes rejections slower, but the architecture of decision trees that have an inbuilt bias to accept, rejecting only at the end of a sequential process.

To test the extent to which these two explanations hold, we conducted a further analysis. First, we confirmed that our experiment replicated Knoch et al.'s (2006) observation that, averaged across all responders, rejections take longer than acceptances (regression of ln(RT): coefficient of rejection dummy = 0.23, $p = 0.000$). We then analyzed a subset of the decisions made by “users” of the three social trees; namely, those decisions made on the basis of the respective final criterion (i.e., the kindness (p.61) criterion or the mirror criterion). Even at this last stage, the conflict between selfish impulses and fairness considerations highlighted by Knoch et al. could be relevant. Indeed, we found that even when the last

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

criterion had to be consulted, rejections took longer than acceptances. The effect size, however, is smaller than for all decisions made using the social trees (regression of ln(RT): coefficient of rejection dummy = 0.14, $p = 0.033$). These results suggest that rejections take longer, not only because selfish impulses have to be overridden (where the proposed payoff exceeds zero), but also because they occur only after one or more other criteria have been consulted.

To conclude, all three response-time predictions received support from the experiment reported here. Responders whose decisions could be modeled in terms of social trees had longer response times than responders whose decisions were better described by the selfish tree (Prediction 1). We also found that as the complexity of the social trees (in terms of looked-up payoffs and mental operations) increased, so did response times (Prediction 2). Finally, we found that the more criteria that a given model predicted would have to be consulted to make a decision, the longer the response times (Prediction 3). This last result is particularly important because it shows that how quickly a given person responds to a proposed division depends in a predictable way on the game under consideration.

Social preference models could be made consistent with Prediction 1 by dropping the inequity parameters (and thus their associated calculations) from the utility function for selfish responders. They could also accommodate Prediction 2 by assuming that other-regarding preferences vary in complexity and that the processing of more complex preferences requires more time. We do not see, however, how further repairs could address the conflict between social preference models and Prediction 3, which states that the response time of the very same responder changes as a function of the game (i.e., the number of criteria that must be retrieved). As long as social preference models evaluate all potentially relevant terms in a given person's utility function, they cannot explain the observed dependency of response times on games.

Let us reiterate: We tailored the classification trees to the choices made in mini-ultimatum games by responders

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

subsumed under Models 1 through 4. That is, we fitted the trees to the behavioral outcomes post hoc. We then derived response predictions from the trees and tested those predictions. The results suggest that a substantial proportion of respondents take various social considerations into account and that those whose decisions can be described in terms of the priority heuristic or the priority+1 heuristic process those considerations sequentially. We now discuss some issues emerging from these results.

(p.62) How Robust Is the Modeling?

Using a mixture-model analysis, we identified four models of responders and derived four classification trees. How robust are these models and trees? As our data set is too small to cross-validate systematically, we must leave it to future investigations to address questions about the extent to which the classification trees generalize to other people and to other mini-ultimatum games. Robustness is as much of an issue for social preference models, however, as for classification trees. Largely neglected in the past, it has recently begun to be raised in the economic literature. For instance, Blanco, Engelmann, and Normann (2011) questioned the robustness of Fehr and Schmidt's (1999) model of inequity aversion. Specifically, they elicited parameters of inequity aversion using different games (e.g., ultimatum game, dictator game, public-good game) but the same sample of participants. They found that although the inequality-aversion parameters had predictive power at the aggregate level, at the individual level the correlations between parameter values were low.

Heterogeneity, Behavioral Consistency, and Strategy Selection

Despite their theoretical differences, social preference models and the classification tree approach struggle to address the same question: Why do we observe heterogeneity in social games? To appreciate the difficulties, consider first the nature of the social motives postulated by social preference models. These motives are typically treated as if they were personality traits or dispositions, which are invoked to explain what people do and why people differ in what they do: “Peggy

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

rejected an unequal division because she dislikes inequity, and Carl did not reject it because he does not mind inequity." According to trait-based explanations, if a person has a specific trait, he can be expected to display trait-relevant behavior under diverse conditions relevant to the trait. Indeed, one could argue that social preference models stand squarely in the tradition of the trait school in psychology and philosophy (see Doris, 2002; for a current example of the trait school, see Hill & Roberts, 2010), whose pitfalls have not received much scrutiny in economics. In heated debates in social and personality psychology dating back to the 1960s, trait theories have been fiercely criticized because of their circularity as well as their neglect of situational (environmental) differences as the primary source of behavioral heterogeneity both within and among people (Doris, 2002).

Social preference models aim to steer clear of circularity by assuming that a postulated preference is stable across situations, (p.63) allowing the models to make behavioral predictions for any game; that is, for any choice situation that is specified and determined in terms of (monetary) incentives. This generalizability, in turn, makes it possible to test and cross-validate a given preference theory in novel situations and thus to inoculate it against accusations of circularity. The stability assumption, however, brings to the fore the very issue with which trait models in psychology have struggled, namely, behavioral consistency—or, more accurately, inconsistency—across trait-relevant situations or, in the present case, across games. The first to investigate this issue in the case of social preference models was Levine (1998), who employed the same parameter distribution (measuring altruism) to successfully predict choices across a small set of games. But as mentioned in our discussion of robustness, Blanco et al. (2011) did not find evidence for behavioral consistency in individuals' inequality aversion across a set of four social games. It seems fair to say that, so far, no social preference theory has met the aspiration to predict individuals' choices correctly across a larger number of games.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

In the framework of classification trees and of boundedly rational heuristics more generally (Gigerenzer et al., 1999; Gigerenzer, Hertwig, & Pachur, 2011), the issue of heterogeneity translates into the questions of why the same person uses different heuristics in different situations and why different people use different heuristics in the same situation. Traits are no more the answer to this thorny problem for heuristics than they are for social preference models. Use of different strategies by different people has been attributed to, among other factors, different goals (e.g., speed vs. accuracy; Payne et al., 1993); differences in the statistical structure of social or physical environments (e.g., Brandstätter et al., 2006, 2008; Dieckmann & Rieskamp, 2007; Payne et al., 1993); individual learning by feedback (e.g., Rieskamp & Otto, 2006); differences in knowledge that trigger different strategies (e.g., Schooler & Hertwig, 2005); and differences in expertise (e.g., Garcia-Retamero & Dhami, 2009). A key task for the future surely will be to examine which of these and other factors (e.g., social norms; see Lévy-Garboua, Meidinger, & Rapoport, 2006) can predict which heuristic (or tree) a person will use.

How Are Social Preference Models and Classification Trees Related?

Our goal was to start developing a new way of modeling behavior in social games that could serve as an alternative to the social preference models proposed in classical economic theory. But does the framework of boundedly rational heuristics in general and of fast and frugal classification trees in particular represent a radical departure (p.64) from social preference models? The authors of this chapter—two economists and one psychologist—do not fully agree on the nature of the relationship between the frameworks. We see two possible views.

According to one view, classification trees are psychologically plausible implementations of the key selfish and social motives postulated by social preference models. Economic models traditionally focus on the behavioral outcome—in the present case, the decision to accept or reject a specific division of a

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

monetary pie—without aspiring to capture the cognitive or affective processes that produce the decision. In this sense, they are *as-if* models or, to use Hoffman's (1960) time-honored term, “paramorphic representations” of people's decisions. Economic models do not pretend that the psychological processes underlying decisions are akin to calculating the equations that form these models’ core. Indeed, models of heuristics complement utility models insofar as they describe how people may arrive at the decision without going through the complex utility calculations implied by, say, theories of inequity aversion by taking advantage of the principles of bounded rationality, such as ordered and limited search and one-reason decision making (chapter 1). On this view, the heuristics are ancillary to the utility models in that they do not render choice predictions that differ from those of utility models, precisely because their sole task is to translate decisions derived from utility calculations into psychological processes.

According to the other view, the two classes of models are competitors on a par in the marketplace of theories. In contrast to *as-if* models, models of heuristics are intended to predict decisions and capture the underlying psychological processes by, for instance, specifying the order of criteria, a stopping rule, and a decision rule. As a consequence, these models can be tested at two levels: decisions and processes. For instance, if a given heuristic predicts decisions as well as social preference models, it may still fail to describe the process, making it false as a process model yet valid as a decision model. When models that rely on highly different theoretical concepts—e.g., utility calculations versus sequential search—fare equally well in predicting decisions (and are equally complex; Roberts & Pashler, 2000), other benchmarks are needed to distinguish between them. Tests of process predictions can act as such benchmarks. If models of heuristics predict outcomes as well as utility models and make accurate process predictions, too, then they will have an explanatory edge over utility models.

Finally, on the second view, models of heuristics can even arrive at predictions that are distinct from those of social preference models. For illustration, take the priority tree.

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Instead of explaining (p.65) rejections as the desire to punish an inequitable division (as inequity-aversion models do) or attributing rejections to a desire to punish unkind behavior (as reciprocity models do), it combines the two concerns into a single tree structure. It can therefore arrive at distinct predictions. Because the priority tree ranks inequity aversion above the kindness criterion, it also gives rise to clear-cut predictions about response time, whereas by their nature social preference models do not.

Psychological Realism

Perhaps more than anyone in economics, Herbert Simon stressed that individual decision makers have no choice but to make decisions under the constraints of limited cognitive resources (e.g., Simon, 1978). On the basis of this reality about the human mind, he criticized classical economic theory for postulating an omniscient rationality that assumes unbounded knowledge, computational capacities, and time. He also targeted Milton Friedman's (1953b) famous defense of it. In response, Friedman argued, “Complete ‘realism’ is clearly unattainable, and the question whether a theory is realistic ‘enough’ can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand” (p. 41).

In Friedman's (1953b) view, the purpose in hand is to account for aggregate behavior; that is, the behavior of firms, institutions, or, more generally, the market. Therefore, unrealistic assumptions and possible discrepancies between theoretical predictions and individual choice behavior need not undermine economic theory. Since Friedman's time, what qualifies as realism in economic models has evolved, as evidenced by, for instance, the rapid rise of neuroeconomics and behavioral economics. There are different ways of opening the black box of decision making, and reasonable people can disagree over whether process models of heuristics or psychological utility models are key to opening it (for one perspective, see Berg & Gigerenzer, 2010). The authors of this chapter, however, are united in their belief that striving for a better understanding of the covert processes of the human mind will lead to better models of human behavior. (p.66)

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017

Access brought to you by: Max-Planck Society

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2017. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see http://www.oxfordscholarship.com/page/privacy-policy). Subscriber: Max-Planck Society; date: 22 February 2017