Rationalization[*]

Vadim Cherepanov, Timothy Feddersen and Alvaro Sandroni.

August 25, 2008

Abstract

In 1908 the Welsh neurologist and psychoanlayst Ernest Jones described human beings as *rationalizers* whose behavior is governed by "the necessity of providing an explanation." We construct a formal model of rationalization. In our model a decision maker is constrained to select the best feasible alternative (according to her preferences) from among those that she can rationalize. We show that this theory is falsifiable and can be tested non-parameterically like the standard theory of choice. We also show that the theory of rationalization subsumes the standard theory and several alternative theories. Rationalization theory can accommodate behavioral patterns often presented in the empirical literature as *anomalies* (i.e., violations of the standard theory of choice). Hence, these anomalies are consistent with the basic principle in economics that choice follows from a constrained optimization process. Moreover, anomalies like cylic choices do not imply cyclic preferences and can be accommodated with preference orders. In fact, anomalies can be used to make inferences about the decision maker's preferences. Rationalization theory reveals a unique preference order in a variety of cases when standard theory cannot. Conversely, when standard theory reveals an order rationalization reveals the same order. These results show that, under suitable assumptions, rationalization theory allows for complete, non-parametric identification of preferences. In addition, rationalization theory can be easily incorporated into game theory.

[*] We thank Eddie Dekel, Jennifer Jordan, Paola Manzini, Marco Mariotti, Scott Presti, Marciano Siniscalchi for useful comments. Vadim Cherepanov [vadimch@sas.upenn.edu] and Alvaro Sandroni [sandroni@sas.upenn.edu] may be reached at the Department of Economics, University of Pennsylvania. Tim Feddersen [tfed@kellogg.northwestern.edu] may be reached at the Kellogg School of Management, Northwestern University.

1. Introduction

In 1908 the Welsh neurologist and psychoanalyst Ernest Jones wrote a paper entitled "Rationalisation in Every-day Life." Jones writes: "[e]veryone feels that as a rational creature he must be able to give a connected, logical and continuous account of himself, his conduct and opinions, and all his mental processes are unconsciously manipulated and revised to that end." While Jones credits Freud with the critical insight "that a number of mental processes owe their origin to causes unknown to and unsuspected by the individual" his paper provides a careful definition of the process of *rationalization–"the necessity of providing an explanation."*

The idea of rationalization has become so well internalized that pundits write about it in the popular press.[1] Psychologists emphasize the facility with which people seemingly create explanations for their behavior. While some explanations may be colorful, the need to rationalize does not have an economic impact unless, on occasion, the inability to rationalize a preferred choice constrains behavior.

We develop a formal model of rationalization. Like in the standard theory, a decision maker (Dee) has preferences over alternatives but, unlike the standard theory, Dee also has a set of *rationales* (modeled as a set of binary relations). Dee chooses the alternative she prefers from among the feasible options she can *rationalize* i.e., those that are optimal according to at least one of her rationales. Thus, rationalization is essentially an exercise in constrained optimization.

Consider the following scenario. Dee decides to take time off from work to see a movie. However, prior to leaving the office she is informed that a colleague is in the local hospital and can accept visitors that afternoon. Dee reconsiders her decision to go to the movie and, instead, stays at work.

Suppose that x denotes attending the movie, y denotes staying at work and z denotes visiting the colleague at the hospital. Dee's behavior can be produced by our model of rationalization as follows. Dee prefers x to y to z (i.e., she prefers the movie to work and work to visiting the hospital). Dee has two rationales available. Under rationale 1 Dee's work is pressing so y is ranked above x and z. Under rationale 2 work is not pressing but Dee must visit the hospital rather than go to the movie. So, rationale 2 ranks z above x and x above y. In a binary choice between x and y Dee chooses x because she prefers x to y and can rationalize this choice using rationale 2. However, if Dee must choose between x, y and z

[1] David Brooks (2008) writes in the New York Times: "In reality, we voters — all of us — make emotional, intuitive decisions about who we prefer, and then come up with post-hoc rationalizations to explain the choices that were already made beneath conscious awareness."

she chooses y because she can't rationalize her most preferred option but can rationalize her second best choice using rationale 1.

Dee's behavior in the example above cannot be accommodated by standard economic theory. As is well-known, observed choice is consistent with the standard theory if and only if *the Weak Axiom of Revealed Preferences (WARP)* holds (see Samuelson (1938a) and (1938b)).[2] WARP states that if x and y are the available choices and x is chosen then the decision-maker does not choose y when x (along with, perhaps, other options) is available.

The standard model of choice has produced great advances in economic theory. However, the limitations of the standard theory have become increasingly clear. A variety of behavioral *anomalies* (i.e., violations of WARP) have been documented in field and laboratory experiments. These anomalies are not mere curiosities, but seem central to understanding behavior in areas such as contribution to public goods, voting, and marketing. We show that rationalization theory can accommodate a wide variety of behavioral anomalies.

Indeed, at first blush, it may seem that any choice behavior whatsoever can be rationalized. However, we show that observed choice is consistent with rationalization theory if and only if *the Weak Weak Axiom of Revealed Preferences (WWARP)* holds. *WWARP* is a familiar relaxation of WARP (see Manzini and Mariotti (2007a)). It requires that if (a) x is chosen over y when x and y are the only alternatives and (b) x is chosen over y from a larger set of alternatives B containing both x and y then y is not chosen from any subset of B that also contains x. This result shows that rationalization theory has empirical content and can be tested in a non-parametric way akin to the standard theory.

In section 1.1, we show several well-known behavioral anomalies can be accommodated by rationalization theory. So, these anomalies are consistent with the basic principle in economics that agents behave as if they (constrained) optimize a single stable, well-defined preference relation. Moreover, Dee's preference relation can be a regular order even if the observed choices are anomalous (e.g., a cycle).

The standard theory is a special case of rationalization theory. In addition, rationalization theory subsumes a variety of alternative theories of choice including warm-glow giving (see Becker (1974) and Andreoni (2006)), rational short-list method (see Manzini and Mariotti (2007a)) and post-dominance rationality theory (see Rubinstein and Salant (2006)). This results show the broad scope of rationalization theory.

We now turn to the matter of inferring preferences from behavior. When Dee

[2] We use the acronym WARP to replace "the Weak Axiom of Revealed Preference."

chooses x over y, standard theory infers that Dee prefers x to y. This inference is allowed by rationalization theory, but other inferences are also possible (e.g., perhaps Dee prefers y to x but can only rationalize x). Still, under rationalization theory, anomalies reveal, to an observer, part of Dee's preferences. Moreover, for many choice functions, Dee's preferences are completely identified by anomalies.

To broaden the scope of choice functions amenable to complete identification of preferences, we return to Jones who remarked on the ability of people to create explanations for behavior. We incorporate this idea by focusing on sets of rationales that impose the minimal constraints required to accommodate the observed choices. Under this principle, Dee's preference order is completely identified for all acyclic rationalizable choice functions and for several cyclic choice functions as well. Hence, the procedure for recovering preferences is basically different for cyclic and acyclic (anomalous or not) choice functions. Moreover, when observed choice does not involve anomalies, the revealed preference is identical to the preference revealed by standard theory. These results show that rationalization theory delivers non-parametric identification of preferences and, therefore, it expands the domain of observed choices for which welfare comparisons are possible.

We provide some directions for future work in a section that includes a simple application of rationalization theory to games. Because rationalization theory assumes Dee has unique preferences, the payoff matrix can be left intact. Players optimize given the additional constraints imposed by their rationales. We show that rationalization theory can accommodate seemingly non self-interested behavior in strategic settings.

The paper is organized as follows: In subsection 1.1 we provide a typology of several behavioral anomalies presented in the empirical literature and show that rationalization can accommodate them. In subsection 1.2 we provide a brief literature review. In section 2, we develop the theory of rationalization formally. Section 3 shows that this theory of choice is fully characterized by WWARP and can subsume several alternative theories. In section 4, we show how anomalies deliver inferences of preferences. In section 5, variations of the basic theory are presented. In section 6, we show how rationalization theory can used for complete, non-parametric identification of preferences. In section 7, we discuss how to empirically reject our theory. In section 8, we discuss directions for future work such as how to incorporate rationalization theory into game theory. Section 9 concludes. All proofs are in the appendix which also contains formalizations of alternative models of choice.

1.1. Anomalies

Given a set of alternatives B, let $C(B)$ be Dee's choice. With three alternatives (say x, y and z), behavioral anomalies can be broken down into three types: *cycles*, *attraction effects* and *difficult choices*.[3]

In a *cycle*,

$$C(\{x,y\}) = x;\ C(\{y,z\}) = y;\ \text{and } C(\{x,z\}) = z.$$

In an *attraction effect*,

$$C(\{x,y\}) = x;\ C(\{y,z\}) = y;\ C(\{x,z\}) = x \text{ and } C(\{x,y,z\}) = y.$$

In a *difficult choice*,

$$C(\{y,z\}) = y;\ C(\{x,z\}) = x \text{ and } C(\{x,y,z\}) = z.$$

All three anomalies were repeatedly observed in economically relevant field and laboratory experiments. The nomenclature follows the empirical literature.

Cycles have been noted at least since May (1954). Manzini and Mariotti (2007a) provide a review of the empirical literature on cycles. Consider the discrimination study of Snyder, Kelck, Stretna and Mentzer (1979). They observe that some subjects watch movie 1 alone (option x) over watching movie 2 alone (y). However, subjects watch movie 2 alone over watching movie 1 with a person in a wheelchair (z) and watch movie 1 with a person in a wheelchair over watching movie 1 alone. So, x is chosen over y, y is chosen over z and z is chosen over x.

Rationalization theory can accommodate the behavior in Snyder et al as follows. Dee prefers x over y over z. All her rationales place z over x and some rationales place x over y and y over z. Informally, Dee prefers to not see the movie with the person in a wheelchair but can only rationalize this choice when the movies are different.

In the attraction effect an alternative (say z) is not chosen but alters choice. When z is not available Dee chooses x over y, but when z is available Dee chooses y over x and z. Ok et al (2008) provides a survey on the empirical evidence for the attraction effect and develops a theory that can accommodate it.[4] Consider the

[3] Let y be the choice from x, y and z and fix the choice between x and z. So, with three alternatives, there are four distinct behavioral patterns (given by the choices on $\{x,y\}$ and $\{z,y\}$). One of them is consistent with WARP and the other three patterns are distinct anomalies.

[4] See also Masatlioglu and Nakajima (2007), Masatlioglu and Ok (2007), Eliaz and Ok (2006).

contribution to public goods study of Berger and Smith (1997). They find that some potential donors (to universities) elect to make a small solicited contribution (x) over no contribution (y), but if either a small or a large contribution (z) is solicited then many people do not contribute at all. Rationalization theory can produce this anomaly as follows: Dee prefers to make a small donation over not donating anything. Her least preferred choice is a large donation. Dee has two rationales. Rationale 1 places y over x over z. Rationale 2 places z over x over y. So, Dee chooses a small contribution over no contribution because she prefers it and can rationalize this choice. Between all three alternatives she selects to not contribute (y) because she cannot rationalize her first choice. Informally, Dee cannot rationalize a small donation when a large donation is also requested. Note that, in this model, Dee chooses x over z and y over z (i.e., the large donation is not chosen) making this pattern an attraction effect.

In a difficult choice, an alternative (z) is not chosen in pairwise choices with x and y, but z is chosen when all three choices are available. This anomaly was observed by Tversky and Shafir (1992) in several laboratory experiments (see also Simonson (1989), Simonson and Tversky (1992) and (1993)). In a different field experiment in marketing, Iyengar and Lepper (2000) show that the fraction of customers who bought a gourmet jam was significantly larger when presented with a limited selection than when presented with an extensive selection.

The difficult choice anomaly can also be accommodated as rationalization. Let x and y be two kinds of jam. Dee prefers both x and y to not buying any jam (let z be not buying jam). Dee has four rationales. Under rationale 1 z is preferred to x or y; under rationale 2 x is preferred to z; under rationale 3 y is preferred to z, under rationale four x is preferred to y. Informally, Dee cannot rationalize one type of jam over another, unless feasibility requires her to buy jam.

The three-alternative anomalies described above are well-known. Naturally, rationalization theory can also accommodate observed anomalies involving four or more alternatives. For example, Tversky and Shafir (1992) provide a five option anomaly that satisfies WWARP.

Several post-hoc explanations have been offered for these anomalies (including an informal version of rationalization). For example, Shafir, Simonson, and Tversky (1993) argue that "decision makers often seek and construct reasons in order to resolve the conflict and justify their choice, to themselves and to others." They show that anomalies are observed more frequently when subjects are required to explain their choices to a third party. In contrast, our effort is to construct a formal, testable model of rationalization.

1.2. Related Literature

A growing literature focuses on accommodating anomalies as a product of internal conflicts. Kalai, Rubinstein and Spiegler (2002) consider a basic model of multiple selves, where choice is optimal according to one of the selves. Green and Hojman (2007) develop a multiple-self model that has no empirical content, but allows partial inferences of preferences. A literature review on multiple-self models can be found in Ambrus and Rozen (2008) who also develop a multiple-self model.

Manzini and Mariotti (2007a, 2007b) define the WWARP axiom and develop a model of choice that can accommodate cycles. They also find empirical support for WWARP in a set of novel experiments.

In our approach, Dee has several rationales, but only one preference relation. This is significant for many reasons. First, it is straightforward to incorporate rationalization into economic analysis because payoffs need not reflect multiple objectives. Moreover, welfare analysis is transparent because Dee has a unique objective function. Finally, Dee's unique preferences may be revealed from choices.

The word "rationalizability" is used in game theory (see Bernheim (1984), Pearce (1984)).[5] Informally, a strategy can be rationalized if some belief about the play of others make that strategy a best-response. Our model and the Bernheim and Pearce notion of rationalizability are very different but share a common idea that actions can be taken when justified by an argument.

Psychologists who study cognitive dissonance also use the word "rationalization" differently from us. Their basic claim is that people devalue rejected choices and upgrade chosen ones (see Chen (2008) for a review). In the area of motivated cognition, Von Hippel (2005) provides a survey on self-serving biased information processing (see also Akerlof and Dickens (1982), Rabin (1995), Dahl and Ransom (1999), Carrillo and Mariotti (2000) and Bénabou and Tirole (2002)).

A large literature deals informally with rationalization in political science. For example, Achen and Bartels (2006) argue that voters justify their support for candidates by discounting unfavorable data. Mendelberg (2001) claims that policy arguments effectively allow voters to rationalize racial prejudice.

An early literature review on psychology and economics can be found in Simon (1959). More recently, Rabin (1998) and Mullainathan and Thaler (2000) provide surveys on models that explain violations of expected utility theory.[6]

[5] See also Sprumont (2000) for another use of the word "rationalizable" in game theory.

[6] See also Gul and Pesendorfer (2005) for a critique of neuroeconomics that raises broader concerns about integrating psychological models of decision-making into economics.

2. Formal Theory of Rationalization

Let A be a finite set of alternatives. A non-empty subset $B \subseteq A$ of alternatives is called an *issue*. Let $\mathcal{B}$ be the set of all issues. A *choice function* is a mapping $C : \mathcal{B} \longrightarrow A$ such that $C(B) \in B$ for every $B \in \mathcal{B}$. Hence, a choice function takes an issue as input and returns a feasible alternative (i.e., the choice) as output.

A binary relation R on A is called a *preference* if it is asymmetric. A transitive, complete preference is an *order*. By standard convention, $x \; R \; y$ denotes that, x is $R-$preferred to y. An alternative $x \in B$ $R-$*optimizes* B if $x \; R \; y$ for every $y \in B$, $y \neq x$. So, x $R-$*optimizes* B if x is $R-$preferred to any other feasible alternative. By convention x $R-$optimizes $\{x\}$.

We consider a decision maker (Dee) endowed with a set of binary relations $\mathcal{R} = \{R_i, \; i = 1, ..., n\}$. Given an issue B, an alternative $x \in B$ is *rationalized* by $R_i \in \mathcal{R}$ if x R_i-optimizes B. So, $x \in B$ is rationalizable if some binary relation in $\mathcal{R}$ rationalizes it. Let $B_{\mathcal{R}} \subseteq B$ be the sub-issue of rationalizable alternatives.

Dee is also endowed with a preference $\bar{R}$ called Dee's preferences. Dee's choice for a given issue B is the alternative $x \in B_{\mathcal{R}}$ that she prefers (given her preference $\bar{R}$) from among all feasible alternatives that she can rationalize. Formally,

Definition 1. *A choice function C is* rationalized *if there exists a preference $\bar{R}$ and a set of rationales $\mathcal{R}$ such that for any issue $B \in \mathcal{B}$, $C(B)$ $\bar{R}-$optimizes $B_{\mathcal{R}}$.*

The model as presented now is quite unstructured. It is sometimes convenient to work with this general model, but often more structure is useful. Let C be a $*-$rationalized choice function if C is rationalized and the binary choices in $\mathcal{R}$ are transitive and asymmetric. We show (claim 1 in the appendix) that C is $*-$rationalized choice function if and only if it is a rationalized choice function. So, without loss of generality, we henceforth assume that the elements in $\mathcal{R}$ are transitive and asymmetric. To simplify the language, we refer to the binary relation in $\mathcal{R}$ as Dee's *rationales.* We also note that if Dee's rationales are required to be orders then the difficult choice anomaly cannot be accommodated.[7]

Analogously, more structure can be imposed on Dee's preference relation $\bar{R}$. In section 5, we further restrict $\bar{R}$ to be an order and show that several anomalies can be accommodated even if Dee's preference relation is an order.[8]

[7] Several well-know principles, such as Pareto efficiency, do not lead to complete relations.

[8] In economics, agents are often endowed with orders, but there is also a tradition that does not require completeness and transitivity (see, among others, Mas-Colell (1974), Gale and Mas-Colell (1975), Shafer and Sonnenschein (1975), Fishburn and LaValle (1988), for early work).

3. Empirical Content

In this section, we characterize the observable implications of rationalized choice functions. We start by recalling the standard theory of choice and its testable implications. In standard theory, all issues are resolved by a single order. That is, there exists an order $\tilde{R}$ such that for any issue $B \in \mathcal{B}$, $C(B)$ $\tilde{R}-$*optimizes* B. As is well known, the standard theory of choice holds if and only if WARP holds.

WARP A choice function C satisfies the weak axiom of revealed preferences iff

$$x \neq y,\ \{x, y\} \subseteq B,\ C(\{x, y\}) = x \text{ then } C(B) \neq y.$$

WARP states that if x is chosen over y then y is never chosen in the presence of x. We now consider a natural weakening of WARP.

WWARP A choice function C satisfies the weak weak axiom of revealed preferences iff

$$x \neq y,\ \{x, y\} \subseteq B_1 \subseteq B_2,\ C(\{x, y\}) = C(B_2) = x \text{ then } C(B_1) \neq y.$$

WWARP states that if x is chosen over y (when x and y are the only choices) and x is also chosen over y (when x, y and a set B of feasible options) then y is not chosen when x, y and a strict subset of B are the feasible alternatives.

So, WARP states that y cannot be chosen in an issue B_1 that contains x and y if (i) x is chosen over y in a binary choice. WWARP states that y cannot be chosen in an issue B_1 that contains x and y if (like WARP) (i) x chosen over y in a binary choice *and* (ii) x chosen in an issue B_2 that contains B_1. We show that WWARP characterizes the observable implications of rationalized choice functions.

Proposition 1. *A choice function C is a rationalized choice function if and only if it satisfies WWARP.*

Proposition 1 demarcates the scope of rationalization theory. It shows the choice functions can and cannot be accommodated as rationalization. This result also shows that even though Dee's choice is subject to unobservable constraints and few restrictions are placed on her preferences and rationales, observable behavior is not arbitrary and the theory can be non-parametrically tested. Moreover, the empirical content of standard theory of choice and rationalization theory are each based on a single axiom (WARP and WWARP) and these two axioms are directly related to each other: WWARP is a familiar and natural relaxation of WARP (see Manzini and Mariotti (2007a)).

As noted in section 1.1, WWARP can accommodate cycles, the attraction effect and the difficult choice. By proposition 1, rationalization theory can accommodate these anomalies. Hence, all these anomalies are consistent with the basic principle in economics that Dee behaves as if she follows a constrained optimization process of a single stable preference relation.

An intuition for proposition 1 is as follows: Assume rationalized choices and consider options x and y. Suppose that x is chosen when only x and y are feasible and x is chosen when a larger set of feasible alternatives B_2 contains x and y. So, x is rationalizable in B_2. If Dee prefers x over y then Dee does not choose y in any issue that contains x and is contained in B_2. If Dee prefers y over x then none of her rationales rank y above x (otherwise y is chosen over x in the binary choice). Hence, y can not be rationalized (or chosen) when x is available.

The proof of sufficiency is constructive. If WARP holds then many constructions are possible. Dee's preferences may be as in the standard model and she has all possible rationales. Alternatively, Dee has only one rationale determined as in the standard model. If WARP does not hold then, as we show in proposition 2, anomalies reveal part of Dee's preferences. So, we defer the discussion of the construction of preferences and rationales until we get to proposition 2.

The theory of rationalized choice, although falsifiable, can accommodate a wide range of behavioral patterns. For all of these patterns of behavior, rationalization theory delivers an interpretation. However, alternative theories may also accommodate some of these behavior patterns and offer a different perspective on choice. Hence, it is useful to relate different theories of behavior. In the next subsection we show several alternative models of choice that are subsumed by rationalization theory.

3.1. Subsuming Alternative Theories

Roth (2007) lists several examples of practices that could produce gains from trade, but were deemed repugnant and banned. One example is the human consumption of horse meat (illegal in California) and the ban in France on dwarf tossing, in spite of the opposition by dwarfs who were paid for being tossed. Other examples include charging interest on loans (usury), profiteering after disasters, selling pollutions permits, and the commercialization of human organs.

Repugnance can be understood as a psychological constraint. For example, Dee may consider some alternatives repugnant and others non-repugnant. All her rationales in $\mathcal{R}$ rank non-repugnant alternatives above repugnant ones. As a result she never chooses repugnant alternatives (even if she prefers them).

Other theories can also be subsumed by rationalization. Manzini and Mariotti (2007a) develop a model of a decision maker who sequentially applies a sequence of criteria to make decisions. In the first stage, Dee eliminates all alternatives that are not maximal according to a preference R_1. In the second stage, Dee chooses optimally according to a preference R_2 from the alternatives not eliminated by R_1. For example, Dee first eliminates all alternatives that are Pareto dominated and then selects among the remaining alternatives by another criteria such as fairness. Manzini and Mariotti show that the rational short list method holds if and only if WWARP and another axiom, called expansion, hold. The expansion axiom rules out the attraction effect and difficult choice (but not cycles). This result, combined with proposition 1, show that rationalization subsumes rational short list methods.

In a recent follow-up paper Manzini and Mariotti (2007b) develop a variant of basic model called the Rational Shortlist Method by Categorization.[9] This model assumes that a decision maker has a set of categories (each category partitions the set of alternatives) which are applied in sequence. For example, Dee could first eliminate all cars that aren't red, then eliminate German cars and then eliminate convertibles until a decision is made. They show that a choice function is consistent with this decision making method if and only if it satisfies WWARP. Hence, even though rationalization and the rational short list method by categorization are quite different theories, they have the same empirical content.

Rubinstein and Salant (2006) propose a post-dominance rationality theory of choice. This theory assumes that Dee first eliminates alternatives that are dominated (according to an acyclic relation R). Then, Dee chooses the best alternative according to a relation that is complete and transitive when restricted to the alternatives not eliminated by R. In the appendix, we show that rationalization theory also subsumes post-dominance rationality theory.

Now consider the theory of warm-glow giving. In the appendix we present a formal model of warm-glow preferences, but the basic idea can be seen in the contribution example. Assume that Dee believes that, morally, she ought to contribute the largest amount requested by a charity. If Dee is asked to make a small donation she does it. Now suppose the charity asks her for either a small or a

[9] Their paper and ours were developed independently.

large donation. Now the choice she thinks is morally best (the large donation) is too costly for her and she decides against it. Once Dee rejects this option she forgoes the psychological reward, called the warm-glow payoff, for acting morally (the remaining alternatives do not conform to Dee's binding rule of morality). Then, Dee chooses the best alternative according to her selfish order. She does not donate at all.

Warm-glow theory can accommodate a large body of evidence in areas such as public good provision and voting turnout that are difficult to reconcile with the standard theory of behavior (see Andreoni (2006) and Feddersen (2004) for partial surveys on warm-glow). The formalization of warm-glow theory is of independent interest because of its important role in the understanding of the paradox of voting turnout and crowding-out effects. However, our focus here is on the demonstration (also placed in the appendix) that warm-glow theory, although based on different principles from rationalization theory, is also subsumed by rationalization theory.[10]

The fact that many theories of choice are either subsumed or empirically equivalent to rationalization theory shows a wide scope of the theory (we expand on the desirability of a broad scope in section 6). Naturally, however, not every theory of choice is subsumed by rationalization. Consider the theory of multiple selves by Kalai, Gil and Rubinstein (2002).[11] In this theory, Dee has multiple (say two) selves and each self is modelled by an order. Given an issue, Dee chooses by one of these orders (no more structure is imposed). In the appendix, we show that dual-self theory may not satisfy WWARP.

4. Recovering Preferences

In rationalization theory, Dee may have multiple rationales, but she has a single, well-defined, and stable preference. As in the standard theory of choice, the uniqueness of Dee's preference makes rationalization theory well-suited to the fundamental exercise of inferring preference from observed choice.[12] Hence, it is natural to ask when observed behavior permits an inference about preference.

Recall our example in the introduction. Dee chooses a movie (x) over work

[10] Warm-glow theory accommodates cycles, the attraction effect, but not the difficult choice.

[11] At various degrees, many other theories of choice are related to the idea of multiple selfs. See, among several others, Dekel, Lipman, and Rustichini (2001), Easley and Rustichini (1999), Fudenberg and Levine (2006), Gul and Pesendorfer (2001), (2004), and (2005), Maccheroni, Marinacci and Rustichini (2006).

[12] See Moulin (1985) for results on the standard theory of choice.

(y), but she also chooses work over the movie and the hospital visit. By standard theory, the former choice is an indication that Dee prefers x over y and the latter choice is an indication of the opposite. In contrast, rationalization theory infers that Dee must prefer x over y. If Dee chooses x over y then either Dee prefers x to y or she cannot rationalize y in the presence of x. The latter possibility is ruled out if y is chosen when x is feasible. So, an anomaly allows an observer to determine whether Dee prefers x over y or if she is psychologically constrained to choose x.

To properly formalize this result we need some notation. A pair of issues (B, B^*) $\in \mathcal{B} \times \mathcal{B}$ is *nested* if $B \subseteq B^*$; B is the *sub-issue* and B^* is the *super-issue*. A pair of nested issues (B, B^*) $\in \mathcal{B} \times \mathcal{B}$ is at *variance* if

$$B \subseteq B^*,\ C(B^*) \in B,\ \text{and } C(B) \neq C(B^*).$$

That is, a pair of nested issues (B, B^*) is at variance if the chosen alternative for the super-issue is available for the sub-issue, but is not chosen. A pair of nested issues at variance violates WARP (and is, therefore, an anomaly).

Let C be a choice function. A preference $\bar{R}$ and a set of rationales $\mathcal{R}$ are said to *underlie* C if for any issue $B \in \mathcal{B}$, $C(B)$ $\bar{R}-$optimizes $B_{\mathcal{R}}$. Let $\mathcal{I}^C$ be the set of all preferences $\bar{R}$ such that for some set of rationales $\mathcal{R}$, $(\bar{R}, \mathcal{R})$ underlie C. With some abuse of terminology, a preference $\bar{R} \in \mathcal{I}^C$ is said to underlie C.

Proposition 2. *Let C be a rationalized choice function.*

1. If (B, B^*) is a pair of nested issues at variance such that $C(B) = x$ and $C(B^*) = y$, $x \neq y$, then $x\ \bar{R}\ y$ for every $\bar{R} \in \mathcal{I}^C$. Conversely,

2. If there exists no pair of nested issues at variance such that $C(B) = x$ and $C(B^*) = y$, $x \neq y$, then $y\ \bar{R}\ x$ for some $\bar{R} \in \mathcal{I}^C$.

Proposition 2 shows that, through anomalies, rationalization theory delivers a partial identification of Dee's preferences. For some choice functions with the appropriate combination of anomalies, this suffices for a complete identification of *all* preferences (see the appendix for examples).

The proof of proposition 2 builds upon the construction of preferences and rationales in proposition 1. First define a set of issues beginning with the entire set A and then subtracting the choice from that set to form a subset. A rationale is constructed for each such issue that ranks the observed choice as best among all choices available. A second set of rationales are constructed for nested issues at variance. These rationales rationalize the observed choice. We define the preference relation as the one revealed by anomalies. Finally, with the constructed rationales, this preference relation suffices for a decision on all issues.

Note that if, except in the binary choice between x and y, Dee always chooses y when x is available, rationalization still implies a preference for x over y. This may, at first, seem counterintuitive. However, reconsider the example of Dee choosing between the movie, work and visiting the hospital. Now suppose another option is available: e.g., Dee is invited to meet with an unpleasant relative (w). It is plausible that Dee is no more able to rationalize the movie over the meeting than she is able to rationalize the movie over the hospital. So now, although Dee prefers x over y, there are three issues ($\{x, y, z\}$, $\{x, y, w\}$ and $\{x, y, z, w\}$) in which Dee chooses y when x is available and only one in which Dee chooses x over y.

The basic logic underlying inferring preference under rationalization theory is not statistical and it does not depend upon the number of issues in which x is chosen when y is feasible. Instead, the inference of preference depends upon a basic property of optimization: if an option is rationalizable in an issue then it is rationalizable in all its sub-issues but not necessarily in its super-issues.

We also wish to emphasize a caveat on the exercise of inferring preferences from choices. Dee's preferences can be partially revealed from anomalies under rationalization theory. However, different inferences about preferences may be reached under the lenses of alternative theories, even if these alternatives theories are subsumed by rationalization theory.

Recall the contribution example. Dee makes the small donation when asked. When the choice of a large donation is added, Dee does not donate. This is an anomaly and, under rationalization theory, Dee must prefer to make a small donation rather than to not donate. Under the lenses of warm-glow theory, a different interpretation is possible: By Dee's selfish order, she prefers not to donate. However, Dee's morality directs her to donate as much as possible. So, in warm-glow theory, Dee makes a small donation because she feels that it is the morally right action. When a large donation is also requested, Dee makes no donation because the ethical choice (the large donation) is now too costly. Unlike rationalization theory, warm-glow theory does not strongly support the claim that Dee prefers to make a small donation. One of Dee's orders (her moral order) prefers to make a small donation, but her other order (her selfish order) prefers to not donate. So, even though warm-glow theory is subsumed by rationalization theory, the two theories may interpret Dee's motivations for the same observed choices differently.

5. Preference Orders

So far we have assumed that preferences are asymmetric binary relations. This raises the concern that the ability to accommodate anomalies follows from the lack of structure on preferences. For example, cyclic choice could be produced from cyclic preferences. In this section we fully characterize the special case of rationalization theory in which Dee's preferences are assumed to be orders. We show that cycles, attraction effects and difficult choices can still be accommodated.

Definition 2. *A choice function C is an order rationalized choice function if there exist an order $\bar{R}$ and a set of rationales $\mathcal{R}$ s.t. $\left(\bar{R}, \mathcal{R}\right)$ underlie C.*

The assumption of preference orders permits additional inferences on preferences. Recall that if $(B,\ B^*)$ is a pair of nested issues at variance such that $C(B) = x$ and $C(B^*) = y$, then, by proposition 2, x must be preferred to y. In this case, we say that x is *directly revealed* to be preferred to y. If z_i is directly revealed preferred to z_{i+1}, $i = 1, ..., k-1$, then we say that z_1 is *indirectly revealed* to be preferred to z_k. Let the binary relation R^p be defined by $x\ R^p y$ whenever x is revealed (directly or indirectly) preferred to y. This terminology is justified by proposition 3 below. For an order rationalized choice function C, let $\mathcal{J}^C \subseteq \mathcal{I}^C$ be the set of all orders in $\mathcal{I}^C$.

Proposition 3. *Let C be an order rationalized choice function.*

1. If x is revealed as preferred to y (i.e., $x\ R^p y$), then $x\ \bar{R}\ y$ for every $\bar{R} \in \mathcal{J}^C$.
2. If x is not revealed as preferred to y, then $y\ \bar{R}\ x$ for some $\bar{R} \in \mathcal{J}^C$.

Proposition 3 shows that Dee's preference order can be partially identified either directly or indirectly through anomalies.[13] We now characterize the observable implications that an order rationalizable choice function must satisfy.

Definition 3. *Fix a choice function C. A set $S = (s_1 ... s_m) \subseteq A$ is called a binary chain if $\forall i \in 1...m-1 : s_i \neq s_{i+1}$, and $\exists S_i^* \supseteq S_i \supseteq \{s_i, s_{i+1}\}$ such that $C\left(S_i\right) = s_i$, and $C\left(S_i^*\right) = s_{i+1}$.*

[13] See Bernheim and Rangel (2007) for an approach to welfare economics which do not require preference orders. We also refer the reader to Chambers and Hayashi (2008) for an alternative way of infering utility functions from observed choice.

That is, a binary chain is a sequence of choices $(s_1...s_m)$ such that s_i directly revealed preferred to s_{i+1}. So, s_1 is indirectly revealed preferred to s_m.

NBCC A choice function C satisfies the No Binary Chain Cycles Axiom if there exists no binary chain $S = (s_1...s_m)$ such that $s_1 = s_m$.

Note that the NBCC axiom does not rule out cyclic choice functions (formally, a choice function C is *cyclic* if there are elements x, y and z such that $C\{x, y\} = x$, $C\{y, z\} = z$ and $C\{x, z\} = z$. A choice function C is *acyclic* if it does not produce any cycles). NBCC only rules out choice functions that reveal a cyclic preference relation. For example, the cycle in section 1.1 satisfies NBCC.

Proposition 4. *A choice function is an order rationalized choice function if and only if it satisfies the NBCC axiom.*

Proposition 4 demarcates the scope of order rationalization theory. The choice functions that can be accommodated are those such that the revealed binary relation R^p is acyclic. The cycle, the attraction effect and the difficult choice satisfy NBCC and so, they are all order rationalized choice functions.

The main substantive point here is that the observation of anomalies does not imply that Dee does not have stable preference orders. In particular, a choice cycle does not necessarily reflect cyclic preferences. The same applies to the attraction effect: it can also be accommodated even if Dee's preferences and rationales are restricted to be orders. And, even if Dee's observed behavior is a difficult choice, her preference may still be an order and her rationales asymmetric and transitive (but may not be complete).

By propositions 1 and 4, NBCC implies WWARP. This can also be seen directly. The irreversibility axiom (IA) states that if x is directly revealed to be preferred to y then y cannot be directly revealed to be prefered to x. It is immediate that NBCC implies IA. In the appendix we define IA in terms of observed choice and show that IA is equivalent to WWARP.

6. Identifying Preference from Choice

When observed choice satisfies WARP, standard welfare analysis infers that the binary choices reveals Dee's preferences.[14] However, standard theory cannot reveal a preference when choice violates WARP. This limits the capacity of modelers

[14] The literature on the empirical estimation of utility functions is too large to be summarized here. We refer the reader to Blow, Browning, and Crawford (2008) for a recent contribution.

to employ standard models of choice and also the ability to make welfare comparisons. A basic question is whether a preference might be revealed from observed choice even when violations of WARP are observed. These inferences, if feasible, broadens the scope of possible welfare comparisons.

We now look for a complete identification of preferences for choice functions that may or may be anomalous. A significant feature of our identification process is that it does not make statistical assumptions, nor does it assume that choices were made by random errors or random preference shocks. It is based only on the data and on well-established principles of economics and psychology. Hence, the contribution of this section is a formalization of these principles and a logical exploration of them when combined with rationalization theory.

The economic principle we adopt for preference identification can be seen in a simple example. Assume that Dee must choose between beef and ham and she picks beef. Perhaps she made this choice because she prefers beef over ham. But perhaps she prefers ham and this choice reflects a religious restriction that does not allow the rationalization of her preferred choice. Finally, Dee may have never tried ham, has no preferences over beef and ham, but her religious restrictions dictate her choice.[15] All three explanations for the same observation may be sensible and rationalization theory allows all of them. However, standard economics adopts the philosophical principle that, in the absence of additional information, we must assume that Dee's choice is unconstrained and, hence, her choice reveals her preference.[16] We adopt the same principle here, but only to the extent that it does not contradict the observed choices. That is, we formalize this principle by considering sets of rationales that are as unrestrictive as possible, but can still underlie the observed choices.

Definition 4. *Consider two pairs $(R, \mathcal{R})$ and $(R', \mathcal{R}')$ that underlie a choice function C. The pair $(R, \mathcal{R})$ is dominated by $(R', \mathcal{R}')$ if R' is an order, $\mathcal{R} \subseteq \mathcal{R}'$, and $B_{\mathcal{R}} \subset B_{\mathcal{R}'}$ for some issue $B \in \mathcal{B}$.*

Consider two pairs $(R, \mathcal{R})$ and $(R', \mathcal{R}')$ that underlie a choice function C. By definition, both pairs can accommodate the observed choices C. However, if

[15] It may also be that Dee prefers beef to ham because of a religious restriction on ham, but this is, of course, just a special case of she preferring beef to ham.

[16] The ideas in this principle can be traced back to the libertarian school of taught that opposes paternalistic polices based on welfare criteria that are inconsistent with the notions of well-being upheld by the individuals affected by these policies (see, for example, Mills (1860)). See also Thaler and Sunstein (2003) for a discussion of paternalism and modern behavioral economics.

$\mathcal{R} \subseteq \mathcal{R}'$ then $\mathcal{R}'$ has all the rationales in $\mathcal{R}$ and, hence, the set of rationalized choices is less restrictive under $\mathcal{R}'$ than under $\mathcal{R}$. Formally, $B_{\mathcal{R}} \subseteq B_{\mathcal{R}'}$ for all issues $B \in \mathcal{B}$. If $B_{\mathcal{R}} \subset B_{\mathcal{R}'}$ for some issue $B \in \mathcal{B}$ then the rationales in $\mathcal{R}$ are strictly more restrictive than those in $\mathcal{R}'$. Hence, the domination principle rules out $(R, \mathcal{R})$ because this pair requires more constraints on choice than are needed to accommodate Dee's observed choices, even if Dee's preference, in the alternative pair $(R', \mathcal{R}')$, is restricted to be an order. This leads to the following definition.

Definition 5. *A preference R is recovered by a choice function C and the domination principle if there exists a set of rationales $\mathcal{R}$ such that $(R, \mathcal{R})$ underlie C and $(R, \mathcal{R})$ is not dominated by any other pair $(R', \mathcal{R}')$ that also underlie C. If a unique preference R is recovered then R is said to be completely identified.*

So, the domination principle selects among all possible preferences and rationales, those that are consistent with the observed choice and do not make use of unnecessary constraints.

As noted above, the domination principle is an extension (to rationalization theory) of the basic criteria that is implicit in standard economics. However, as noted in the introduction, the domination principle can also be motivated by the common perception in psychology that people can easily rationalize choice and only on occasion find themselves unable to rationalize their preferred option. The next result follows directly from the finiteness of the set of alternatives.

Proposition 5. *If C is a rationalized choice function then at least one preference is recovered by C and the domination principle. If C is an order rationalized choice function then at least one preference order is recovered by C and the domination principle.*

By proposition 5, the domination principle always permits inferences over Dee's preferences. We now show how this principle may help identify Dee's preferences. Given a choice function C, let $\tilde{R}^C$ be the binary relation such that $x \; \tilde{R}^C \; y$ if and only if $C(\{x, y\}) = x$, $x \neq y$. We call $\tilde{R}^C$ the *binary choice relation* because it is determined entirely by observed binary choices. By construction, $\tilde{R}^C$ is complete and asymmetric. Moreover, $\tilde{R}^C$ is an order if the choice function C is acyclic.

Proposition 6. *Let C be an acyclic, rationalized choice function. Then, the binary choice relation is completely identified by the choice function C and the domination principle.*

Proposition 6 shows that when a rationalized choice function is acyclic, binary choices are revealed as the unique preference order implied by rationalization theory and the domination principle.[17] Among all possible preferences (orders or not), the only surviving one is the order given by the binary choices. So, with the assistance of the domination principle, rationalization theory strictly extends the domain of choice functions from which preferences can be completely identified. Indeed, if the choice function satisfies WARP then it is acyclic and rationalized. In this case, the revealed preference is the same as in standard theory. However, complete identification of preferences now extends to cases where choices may be anomalous such as the attraction effect and the difficult choice.

Consider the following basic question: when is it the case an outsider can infer that x is prefered over y when observing a binary choice of x over y? Proposition 6 delivers a simple and compelling answer. If no cycles are observed then, whether or not Dee's behavior is anomalous, Dee's binary choices reveal her preferences. We now show that this simple procedure for recovering preferences is inadequate for cyclic choice functions. In the presence of cycles, non-binary choice is important for recovering preferences.

Proposition 7. *Assume that C is a rationalized choice function and $\left(\bar{R},\mathcal{R}\right)$ underlie C. Then $\left(\tilde{R}^C,\mathcal{R}\right)$ also underlie C.*

Proposition 7 shows that it is always conceivable that Dee's preferences coincide with her binary choices. Hence, a corollary of propositions 5 and 7 is that if C is a cyclic order rationalized choice function then there is a preference order *and* a cyclic preference that is recovered by C and the domination principle.

Propositions 6 and 7 show a basic difference between cyclic and acyclic anomalies. By proposition 6, when C is acyclic and order rationalized then a unique preference is completely identified. When C is cyclic and order rationalized then at least two preference relations can be recovered.

We now argue that in the case of cyclic choice functions satisfying NBCC preference orders should *not* be ruled out. Recall the cyclic choices described in the introduction. In the binary choices, x (the movie) is chosen over y (to work), y over z (the hospital) and z over x. So, in this example, visiting the hospital is chosen over going out to see a movie, but it seems counterintuitive that Dee prefers the hospital over the movie as this binary choice indicates. Analogously, in

[17] A corollary of proposition 6 is that an acyclic, rationalized choice function is an order rationalized choice function.

the discrimination example described in section 1.1, Dee chooses to see the movie with the handicapped person (z) over seeing the same movie alone (x), but it also seems counterintuitive that this choice represents her preferences given that she chooses to see the movie alone rather than with the handicap when the movies are different. Hence, in both examples, the order x preferred to y and y to z seems intuitive and the cyclic preference x over y, y over z and z over x seems counterintuitive.

The general procedure for recovering orders is as follows: Let C be an order rationalized choice function. If C is acyclic then Dee's order is identified by her binary choice relation. If C is cyclic then first find her revealed acyclic preference determined by the observed anomalies in section 5. Consider all orders that extend this acyclic relation. Any of them underlie C, but only the orders that survive the domination principle are the recovered orders.

Rationalization theory may, by itself and without applying the domination principle, imply a unique order for some cyclic choice. That is, for some cyclic choice functions, only one order underlies it. In other cyclic choice functions, more than one order underlies it, but only one order is recovered when the domination principle is applied. Finally, there are cyclic choice functions with more than one recovered order even under the domination principle (examples of all these cases are provided in the appendix).

Whether it is possible to obtain complete identification of preferences for a broader class of choice functions is an open and important question. In section 8, we consider the idea that rationales are partially social constructions and, hence, there are grounds for the assumption that Dee has some intuitive rationales. In that case a complete identification of preferences becomes possible even in cases where the domination principle does not select a single order.

Note that while complete identification of preferences hold for several, but not all, rationalized choice functions, partial identification of preferences hold for all rationalized choice functions C (under the domination principle). Indeed, if C satisfies WARP then a unique order is identified. If C does not satisfy WARP then partial identification follows from proposition 2.

6.1. Intuition of Propositions 6 and 7

The intuition behind proposition 7 is as follows: consider a preference $\bar{R}$ and set of rationales $\mathcal{R}$ that underlie a choice function C. Now, consider an issue B and a rationalizable choice $z \in B_{\mathcal{R}}$ that differs from the actual choice $C(B)$. Since both

z and $C(B)$ are rationalizable in B and $C(B)$ is the choice in B it follows that $C(B)$ is $\bar{R}-$preferred to z. Next consider the binary choice between $C(B)$ and z. Both $C(B)$ and z must be rationalizable and $C(B)$ must be $\bar{R}-$preferred to z (because $C(B)$ was chosen when both alternatives were rationalizable). So, given that $\left(\bar{R}, \mathcal{R}\right)$ underlie C, it follows that in the binary choice, $C(B)$ is chosen over z. So, $C\left(B\right)$ is also $\tilde{R}^{C}-$preferred over z. Hence, $\left(\tilde{R}^{C}, \mathcal{R}\right)$ underlie C.

The intuition behind proposition 6 is as follows: Note that if $(R,\, \mathcal{R})$ is not dominated then both options x and y can be rationalized in the binary choice. Otherwise consider an alternative set of rationales $\mathcal{R}'$ comprising of all the rationales in $\mathcal{R}$, but and a new rationale that rationalizes both x and y in the binary choice. The set of rationales $\mathcal{R}'$ can be paired with $\tilde{R}^{C}$ to underlie choice, contradicting the assumption that $(R,\, \mathcal{R})$ is not dominated. But if both choices can be rationalized in any binary choice then Dee's preferences must coincide with $\tilde{R}^{C}$.

7. Rejecting the Theory

Manzini and Mariotti (2007b) set up an experiment and show that violations of WWARP, while rare in their experiment, may occur. The simple experiment that follows may also produce violations of WWARP.

Simonson and Tversky (1992) and (1993), see also Hubler et al (1982), observe the following anomaly. Dee chooses cash over an elegant pen, but the elegant pen is selected when a regular pen is added as an option.

A variation of this experiment could have two distinct objects x_1 and x_2 (e.g., x_1 is an elegant pen and x_2 an elegant wallet) and a lower quality variation of the same object y_1 and y_2 (e.g., y_1 is a regular pen and and y_2 a regular wallet). Assume that Dee chooses x_1 over x_2, but selects x_2 when the options are x_1, x_2 and y_2. So far, this is identical to the experiment in Simonson and Tversky (1992), with a wallet replacing cash. However, if Dee chooses x_1 when the options are x_1, x_2, y_1 and y_2 then a violation of WWARP is obtained and the theory of rationalization is rejected.

These choices are plausible. If the lower quality pen enhances the attractiveness of the elegant pen then the lower quality wallet my also enhance the attractiveness of the elegant wallet. If the elegant wallet is chosen over the elegant pen then the elegant wallet may be chosen over the elegant pen, the regular pen and the regular wallet.

8. Directions for Future Work

8.1. Application to Game Theory

Even in the simple form presented here rationalization theory can be incorporated into game theory. Our objective here is only to demonstrate the kinds of insights such a project might deliver. So, consider the prisoner dilemma with players making rationalized choices. The game is

(I, II)	C	D
C	$(-1, -1)$	$(-20, 0)$
D	$(0, -20)$	$(-10, -10)$

So, 1's preferences are

$$(D, C) \succ (C, C) \succ (D, D) \succ (C, D).$$

and 2's preferences are

$$(C, D) \succ (C, C) \succ (D, D) \succ (D, C).$$

The rationales of player 1 are all orders such that $(C, C) \succ (D, C)$. The rationales of player 2 are all orders such that $(C, C) \succ (C, D)$. That is, none of the players can rationalize defection (D) if the other player cooperates (C). So, if 2 cooperates then 1's feasible options are (C, C) and (D, C). In this case, 1 cooperates because this is the only rationalizable option. If 2 defects then 1 can rationalize both options (C, D) or (D, D), and 1 defects because she prefers it. An analogous result holds for player 2. So, the profile in which both players cooperate and the profile in both players defect are equilibrium outcomes.

Now assume that player 1 moves first and plays either aggressive (A) or pleasant (P). Player 2 observes the play of 1 and either reciprocates (R) or does not reciprocate (N). Payoffs for (A, R) , (A, N), (P, R) and (P, N) are $(6, 0)$, $(2, 1)$, $(4, 1)$ and $(3, 2)$, respectively. Player 1's rationales are all orders and players 2's rationales are all orders such that $(P, R) \succ (P, N)$. So, if 1 is pleasant then player 2 can only rationalize reciprocation and, hence, plays R. If 1 plays aggressive then player 2 can rationalize both options and plays N. In the only subgame perfect equilibrium, player 1 is pleasant and 2 reciprocates (see Rabin (1993), Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) for alternative models of reciprocity). We also refer the reader to Spiegler (2002) and (2004) for game-theoretic models where players must justify their chosen actions.

A general introduction of rationalization into game theory would require extending the theory to allow for choice over lotteries. Preference relations in rationalization theory are such that, in principle and with suitable restrictions, they can be represented by an expected utility function. Rationales might also need to be restricted to ensure existence of an equilibrium. This is beyond the scope of this paper.

Many rationales (e.g., Pareto optimal alternatives are best) are incomplete and so our minimal assumption of asymmetry on rationales is justified. However, it would still be worthwhile to characterize the empirical content of rationalization theory under the assumption that both preferences and rationales are orders.

8.2. Partial Knowledge of Rationales

Rationales often seem to be social constructions. Even partial knowledge of rationales can allow an observer to make inferences about preferences and choice that are not possible more generally. Consider the three alternative cycle in the introduction. Suppose that an outside observer knows that Dee can rationalize going to the hospital (z) above either work (y) or the movie (x). Since y is the choice from $\{x, y, z\}$ it follows that both y and z are rationalizable when the issue comprises of all three alternatives. The choice of y in this issue (i.e., $C(\{x, y, z\}) = y$) implies that y must be preferred to z. As before, the anomaly implies that x must be preferred to y and so a preference order is completely identified; Dee prefers the movie to work to the hospital. The conditions under which it is sensible to assume partial knowledge of rationales represents an avenue for future research.

A related extension is to consider endogenizing rationales. Viewing decision situations in isolation it is natural to ask why the set of rationales would not be completely flexible and permit the rationalization of any choice. However, in dynamic settings restrictions on choice might serve as a commitment device. In addition, rationales might be the result of social convention that may evolve to constrain selfish behavior. Changing conventional mores (e.g., with respect to marriage, charging interest, relations between adults and children) seem to play an important role in changing behavior. Rationalization theory provides a natural vehicle for exploring the strategic incentives underlying disputes over social mores and stigmatized behavior.

9. Conclusion

We develop a tractable and testable model in which decisions must be rationalized. This model is fully characterized by a simple axiom and can accommodate several behavioral patterns that are incompatible with the standard theory of choice. Morever, these anomalies can be used for partial identification of preferences. With additional assumptions, rationalization theory delivers a complete identification of preferences for a class of observed choice functions that may or may not be anomalous.

10. Appendix

The appendix is organized as follows: In subsection 10.1, we infer preferences from some choice function. In subsection 10.2, we present a formal theory of warm-glow and we shown that this theory is subsumed by rationalization theory. In subsection 10.3, we show that other theories of choices are also related to rationalization theory. Proofs are in subsection 10.4.

10.1. Examples

Let C^1 be a rationalized choice function defined by

$$C^1(x,y) = C^1(x,z) = C^1(x,w) = x;$$
$$C^1(y,z) = C^1(y,w) = C^1(x,y,z) = C^1(x,y,w) = y;$$
$$C^1(z,w) = C^1(y,z,w) = C^1(x,z,w) = z;\ C^1(x,y,z,w) = w.$$

By proposition 2, the choices in C^1 reveal that x is preferred to y,z,w; y is preferred to z,w and z is preferred to w. So, the order x is preferred to y, y to z,w and z is revealed by the observed choices. Hence, if the observed choices are given by C^1 then rationalization theory completely identifies Dee's preferences.

Now consider the choice function C^2 that is the same as C^1, except that $C^2(x,w) = w$. Note that C^2 is cyclic. Given the choices in C^2, the unique preference order recovered is such that x is preferred to y, y to z, and z to w.

Now consider the following choice function C^3

$$\begin{aligned} C^3(x,y) &= C^3(x,w) = x;\ C^3(x,z) = C^3(x,y,z) = z; \\ C^3(y,z) &= C^3(y,w) = C(x,y,w) = y; \\ C^3(z,w) &= C^3(x,z,w) = C^3(y,z,w) = C(x,y,z,w) = w. \end{aligned}$$

Note that C^3 is cyclic and two preference orders underlie it: $x \ \hat{R} \ y \ \hat{R} \ z \ \hat{R} \ w$ and $x \ \bar{R} \ y \ \bar{R} \ w \ \bar{R} \ z$. However, $\hat{R}$ is ruled out by the domination principle. So, $\bar{R}$ is the unique remaining order.

Recall the cyclic choices described in the introduction. In the binary choices, x is chosen over y, y over z and z over x. The symmetry of these choices make it difficult to reveal preferences. So, we must rely on the choice of y when the issue is $\{x, y, z\}$. This choice creates an asymmetry. Dee is revealed to prefer x over y, but the relative position of z is less clear. Three preference orders underlie this cycle: $x \ \bar{R} \ y \ \bar{R} \ z$; $z \ \hat{R} \ x \ \hat{R} \ y$ or $x \ \dot{R} \ z \ \dot{R} \ y$. The order $\dot{R}$ is ruled out by the domination principle, but $\bar{R}$ and $\hat{R}$ are not.

10.2. Warm-glow Theory of Choice

Warm-glow theory can be formalized as follows: Dee has two orders, the moral order R_m and the selfish order R_s. In addition, Dee is endowed with a limit function $l : A \longrightarrow A$ such that

$$\begin{array}{c} \text{either } l(a) = a \text{ or } a \ R_s \ l(a); \text{ and} \\ \text{if } a' \ R_s \ a \text{ then either } l(a) = l(a') \text{ or } l(a') \ R_s \ l(a). \end{array} \tag{10.1}$$

So, the lower limit of a is either a or ranked (by R_s) below a. In addition, if R_s ranks a' above a then the lower limit of a' is either the same as the lower limit of a or ranked (by R_s) above the lower limit of a.

Given an order R and issue B, let $R(B) \in B$ be the alternative that $R-$optimizes it. That is, $R(B) \ R \ b$, for every $b \in B$, $b \neq R(B)$. The existence and uniqueness of $R(B)$ is assured by the assumption that R is an order. Given an issue B, Dee' moral rule directs her to take alternative $R_m(B)$ while her selfish view directs her to take alternative $R_s(B)$. The alternatives in B that R_s ranks below the lower limit $l(R_s(B))$ are those that are too costly to take. Hence, if $l(R_s(B)) \ R_s \ R_m(B)$ then Dee's rule of morality directs her to take an action $R_m(B)$ that her selfish order, R_s, ranks below the lower limit, $l(R_s(B))$, of Dee's preferred outcome (from her selfish point of view). In this case, we say that for issue $B \in \mathcal{B}$, Dee's moral order R_m *directs her to an excessively costly choice.* On the other hand, if $R_m(B) \ R_s \ l(R_s(B))$ then Dee's rule of morality does not direct her to take an action $R_m(B)$ that is excessively costly in the sense that R_s ranks $R_m(B)$ above the lower limit, $l(R_s(B))$.

The conditions in 10.1 are natural requirements. The limit $l(a)$ of an alternative should not be ranked (by R_s) above a because $l(a)$ marks the least attractive option (by R_s) that Dee can take (if morally directed to so) when a is feasible. In addition, if b is too costly in the presence of a then it should remain too costly when an even better opportunity (a' s.t. $a'\ R_s\ a$) becomes feasible.

Definition 6. *A choice function C is a warm-glow choice function if there exists an order R_m, an order R_s and a limit function l that satisfies 10.1 such that for any issue $B \in \mathcal{B}$,*

$$\begin{gathered} C(B) = R_m(B) \ \ \text{when } R_m(B)\ R_s\ l(R_s(B));\ \text{and} \\ C(B) = R_s(B) \ \ \text{when } l(R_s(B))\ R_s\ R_m(B). \end{gathered}$$

So, C is a warm-glow choice function if Dee chooses according to her moral order whenever it does not direct her to an excessively costly choice. If Dee's moral order directs her to an excessively costly choice then she chooses by her selfish order.

Proposition 8. *Any warm-glow choice function satisfies WWARP.*

It follows from propositions 1 and 8 that warm-glow theory is subsumed by rationalization theory. In addition, it is straightforward to show that some anomalies (like the difficult choice) cannot be accommodated by warm-glow theory (but can be accommodated rationalization theory).

10.3. An Alternative Theory of Choice

Now consider post-dominance rationality theory of choice. Rubinstein and Salant (2006) show that post-dominance rationality holds if and only if it the Exclusion Consistency (EC) axiom. EC states that for every issue B and alternative b, $C\left(B \cup \{b\}\right) \notin \{C\left(B\right), b\} \Rightarrow \forall$ issue S such that $b \in S : C\left(S\right) \neq C\left(B\right)$. We now show that post-dominance rationality theory is also subsumed by rationalization theory.

Proposition 9. *If a choice function satisfies EC then it also satisfies WWARP.*

It follows from propositions 1 and 9 that post-dominance rationality theory is subsumed by rationalization theory. We now consider a dual-self theory. Each self is represented by an order. Given an issue, Dee chooses with either order R_1 or R_2. No other restriction is imposed.

Definition 7. *A choice function C is a dual-self choice function if there are orders R_1 and R_2 such that for any issue $B \in \mathcal{B}$, $C(B) \in \{R_1(B), R_2(B)\}$.*

Remark 1. *Dual-self choice functions do not necessarily satisfy WWARP.*

Consider the following example. There are four alternatives $\{x, y, z,$ and $w\}$. The first order ranks y first followed by x, z, and w. The second order ranks x first followed by y, z, and w. Now consider the sets $B_2 = \{x, y, z, w\}$ and $B_1 = \{x, y, z\}$. Now assume that B_2 and $\{x, y\}$ are resolved by order two. Then, x is chosen. Also assume that B_1 is resolved by order one. Then, y is chosen. This is a violation of WWARP.

10.4. Proofs

Recall that a pair of issues $(B,\ B^*)$ is nested if $B \subseteq B^*$ and a pair of nested issues $(B,\ B^*)$ is at variance if $C(B^*) \in B$ and $C(B) \neq C(B^*)$. So, the chosen alternative for the super-issue is available at the sub-issue, but is not chosen. The choices on two pairs $(B_1,\ B_1^*)$ and $(B_2,\ B_2^*)$ of nested issues are *reversed* if $C(B_1) = C(B_2^*)$ and $C(B_1^*) = C(B_2)$. So, the choice in the sub-issue of one of the pairs is the choice on the super-issue of the other pair.

Definition 8. *A choice function C satisfies the irreversibility axiom if there are no two pairs of nested issues at variance with reversed choices.*

Lemma 1. *The irreversibility axiom holds if and only if WWARP holds.*

Proof: Assume that WWARP does not hold. Then let $x \neq y$, $\{x, y\} \subseteq B \subseteq \bar{B}$ be such that $C(\bar{B}) = C(\{x, y\}) = x$ and $C(B) = y$. Then, $(\{x, y\},\ B)$ is a pair of nested issues at variance and $(B, \bar{B})$ is also a pair of nested issues at variance. But $C(\bar{B}) = C(\{x, y\}) = x$. Hence, $(\{x, y\},\ B)$ and $(B, \bar{B})$ are reversed. Thus, the irreversibility axiom does not hold.

Now assume that the irreversibility axiom does not hold. Consider the two pairs $(B_1,\ B_1^*)$ and $(B_2,\ B_2^*)$ of reversed nested issues at variance. Let $y = C(B_1) = C(B_2^*)$ and $x = C(B_1^*) = C(B_2)$. Then, $x \neq y$, $\{x, y\} \subseteq B_1 \subseteq B_1^*$ and $\{x, y\} \subseteq B_2 \subseteq B_2^*$ ($x \in B_1$ because $x = C(B_1^*) \in B_1$ and $y \in B_1$ because $y = C(B_1) \in B_1$. So, $\{x, y\} \subseteq B_1$. The argument for $\{x, y\} \subseteq B_2$ is analogous). Now assume that $C(\{x, y\}) = x$. Then, $\{x, y\} \subseteq B_1 \subseteq B_1^*$, $C(B_1^*) = x$ and $C(B_1) = y$. So, WWARP does not hold. On the other hand if $C(\{x, y\}) = y$ then $\{x, y\} \subseteq B_2 \subseteq B_2^*$, $C(B_2^*) = y$ and $C(B_2) = x$. Thus, WWARP does not hold.■

10.4.1. Proof of Proposition 1

By lemma 1, to show proposition 1, we need to show that C is a rationalized choice function if and only if the irreversibility axiom holds. Assume that C is a rationalized choice function. Let $(B,\, B^*)$ be a pair of nested issues at variance. Then, $B \subseteq B^*$ implies that $B^*_{\mathcal{R}} \bigcap B \subseteq B_{\mathcal{R}}$. This follows because if $x \in B^*_{\mathcal{R}} \bigcap B$ then $x \in B$ and there exists $R_i \in \mathcal{R}$ such that $x\ R_i\ y$ for every $y \in B^*$, $y \neq x$. Thus, $x \in B$ and $x\ R_i\ y$ for every $y \in B$, $y \neq x$. So, $x \in B_{\mathcal{R}}$.

By definition, $C(B)\ \bar{R}\ y$ for every $y \in B_{\mathcal{R}}$, $y \neq x$ and, therefore, $C(B)\ \bar{R}\ y$ for every $y \in B^*_{\mathcal{R}} \bigcap B$, $y \neq x$. Moreover, $C(B^*) \in B^*_{\mathcal{R}}$ (by definition), $C(B^*) \in B$; $C(B^*) \neq C(B)$ (because $(B,\, B^*)$ is at variance). It follows that $C(B)\ \bar{R}\ C(B^*)$. Hence, if $(B_i,\, B^*_i),\ \ i \in \{1,2\}$, is a pair of nested issues at variance then

$$C(B_i)\ \bar{R}\ C(B^*_i),\ i \in \{1,2\}.$$

Given that $\bar{R}$ is asymmetric it follows that $C(B_1) = C(B^*_2)$ and $C(B^*_1) = C(B_2)$ leads to a contradiction. Hence, the choices on these two pairs cannot be reversed.

Note that the proof that rationalized choice functions satisfies the irreversibility axiom does not require the assumption that Dee's preference relation, $\bar{R}$, are transitive.

Now assume that C satisfies the irreversibility axiom. Let $\bar{R}$ be defined as follows: $x\ \bar{R}\ y$ if and only if there exists a pair of nested issues at variance $(B,\, B^*)$ such that $C(B^*) = y$ and $C(B) = x$. By the irreversibility axiom, $\bar{R}$ is asymmetric.

Let $\bar{\mathcal{B}} \subseteq \mathcal{B}$ comprise of all issues $B \in \mathcal{B}$ such that there exists a super-issue B^* of B such that $(B,\, B^*)$ is a pair nested issues at variance. Assume that A has k elements. We construct, by induction, the following k issues. Let $A_1 = A$. Given an issue $A_j \in \mathcal{B}$, $j < k$, let A_{j+1} be the issue A_j with (only) the alternative $C(A_j)$ removed.

For each $B \in \mathcal{B}$ let R_B be the rationale defined by $C(B)\ R_B\ y$ for every $y \in B$, $y \neq C(B)$. That is, $z\ R_B\ y$ if and only if $y \in B$, $y \neq C(B)$, and $z = C(B)$. By definition, R_B is transitive and asymmetric. Let $\hat{\mathcal{B}} = \bar{\mathcal{B}} \bigcup \{A_j,\ j = 1, ..., k\}$. Let $\mathcal{R} = \{R_B,\ B \in \hat{\mathcal{B}}\}$. We now show that C is a rationalized choice function with $\bar{R}$ as Dee's preference relation and $\mathcal{R}$ as her set of rationales.

We first show that $C(B) \in B_{\mathcal{R}}$. First assume that $B \in \bar{\mathcal{B}}$. Then $B \in \hat{\mathcal{B}}$ and, by definition, $C(B)\ R_B-$optimizes B. Hence, $C(B)$ is rationalized by R_B. Now assume that $B \notin \bar{\mathcal{B}}$. Let $\bar{j} \in \{1, ..., k\}$ be such that $B \subseteq A_{\bar{j}}$ and $C(A_{\bar{j}}) \in B$. Then, $B \notin \bar{\mathcal{B}}$ implies that $C(B) = C(A_{\bar{j}})$. So, $C(B)$ is rationalized by $R_{A_{\bar{j}}}$.

Now assume that $z \in B_{\mathcal{R}}$ and $z \neq C(B)$. Then, $z \in B$ and, for some issue $\mathring{B} \in \hat{\mathcal{B}}$, $z \ R_{\mathring{B}} \ y$ for every $y \in B$, $y \neq z$. Now, $z = C(\mathring{B})$ and $B \subseteq \mathring{B}$ (because if $w \notin \mathring{B}$ then $w \neq z = C(\mathring{B})$ and $C(\mathring{B}) \ R_{\mathring{B}} \ w$ does not hold and so, $w \notin B$). Hence, $\left(B, \mathring{B}\right)$ is a pair of nested issues at variance. Thus, $C(B) \ \bar{R} \ z$. It follows that $C(B) \ \bar{R}-$optimizes $B_{\mathcal{R}}$.■

Remark 2. *Paola Manzini and Marco Mariotti sent us, in private correspondence, a simpler proof of proposition 1. Our construction of preferences and rationales, although more complex, is analytically convenient for other results and so is kept. However, we are grateful to Manzini and Mariotti for their argument.*

10.4.2. Proof of Claim 1

Assume that C is a rationalized choice. Then, by proposition 1, it satisfies WWARP and, by construction, one the pairs $(R, \mathcal{B})$ that underlie C are such that R is asymmetric and all the binary relation in $\mathcal{B}$ are transitive and asymmetric. Hence, C is a $*-$rationalized choice. The converse is immediate.■

10.4.3. Proof of Proposition 2

Assume that C satisfies WWARP, $C(B) = x$, $C(B^{*}) = y$, $\{x, y\} \subseteq B \subseteq B^{*}$, $x \neq y$. Also assume that $(\bar{R}, \mathcal{R})$ underlie C (the existence of $(\bar{R}, \mathcal{R})$ is assured by proposition and the assumption that C satisfies WWARP). Given that $C(B) = x$ it follows that $x \ \bar{R} \ w$ for every $w \in B_{\mathcal{R}}$, $w \neq x$. So, the proof is concluded if $y \in B_{\mathcal{R}}$. Now assume that $y \notin B_{\mathcal{R}}$. This implies that $y \notin B^{*}_{\mathcal{R}}$ (this implication can be shown as follows assume that $w \in B^{*}_{\mathcal{R}}$ then for some $R \in \mathcal{R}$, $w \ R \ z$ for every $z \in B^{*}$, $z \neq w$. So, $w \ R \ z$ for every $z \in B$, $z \neq w$. Hence, $w \in B_{\mathcal{R}}$). It now follows that $C(B^{*}) \neq y$. A contradiction.

Now assume that C satisfies WWARP and there is no pair of nested issues at variance such that $C(B) = x$ and $C(B^{*}) = y$, $x \neq y$. Let $(\bar{R}, \mathcal{R})$ be as constructed in the proof of proposition 1. So, by construction, $(\bar{R}, \mathcal{R})$ underlie C and it is *not* the case that $x \ \bar{R} \ y$. Without loss of generality we can assume that it is also not the case that $y \ \bar{R} \ x$ (otherwise the proof is immediate). Let $\bar{R}'$ be the binary relation that coincides with $\bar{R}$ for all pairs $(w, z) \in A\text{x}A$ such that $(w, z) \neq (x, y)$. In addition, $y \ \bar{R}' \ x$. Clearly, $\bar{R}'$ is asymmetric because $\bar{R}$ is asymmetric and, by construction, $x \ \bar{R}' \ y$ does not hold. Now, we show that $C(B) \ \bar{R}' \ w$, for every

$w \in B_{\mathcal{R}}$, $w \neq C(B)$. By construction, this holds unless $C(B) = y$ and $w = x$. So, assume that $x \in B_{\mathcal{R}}$ and $C(B) = y$. Then, $y\ \bar{R}\ x$. A contradiction. Hence, $C(B)$ $\bar{R}'\ w$, for every $w \in B_{\mathcal{R}}$, $w \neq C(B)$.■

10.4.4. Proof of Propositions 3 and 4

The proof of proposition 4 is as follows. First the direct implication ⇒ . Suppose $S = (s_1 ... s_m)$ is a binary chain. So, $C\left(S_i^*\right) = s_{i+1}$, $s_{i+1} \in S_i$ and $C\left(S_i\right) = s_i$. Now $C\left(S_i^*\right) = s_{i+1}$ implies that for some $R \in \mathcal{R}$, $s_{i+1}\ R\ y$ for every $y \neq s_{i+1}$ such that $y \in S_i^*$. Given that $S_i \subset S_i^*$, $s_{i+1}\ R\ y$ for every $y \neq s_{i+1}$ such that $y \in S_i$. So, $s_{i+1} \in \left(S_i\right)_{\mathcal{R}}$. Hence, $s_i\ \bar{R}\ s_{i+1}$ $i = 1, ..., m-1$. The assumption that $\bar{R}$ is an order now implies that $s_1 \neq s_m$.

Now the converse ⇐ . Suppose now, that choice function C satisfies NBCC. Let us, first, define a relation $\bar{R}'$ on A: $x\bar{R}'y$ if and only if there exists a binary chain $(x...y)$.

a) $\bar{R}'$ is asymmetric: if $x\ \bar{R}'\ y$ and $y\ \bar{R}'\ x$, then there are binary chains $(x...y)$ and $(y...x)$, the union of these two binary chains $(x...y...x)$ is a cyclical binary chain.

b) $\bar{R}'$ is transitive: suppose $x\bar{R}'y$, and $y\bar{R}'z$, then there exist binary chains $(x...y)$ and $(y...z)$, and, hence, their union $(x...y...z)$ is binary chain, so, $x\bar{R}'z$.

By topological ordering, $\bar{R}'$ may be extended (not necessarily uniquely) to an order (see Cormen et al. (2001, pp.549–552)). Let $\bar{R}$ be an arbitrary order that extends $\bar{R}'$. Let $\mathcal{R}$ be as defined in the proof of proposition 1. By the same argument as in the proof of proposition 1, for each $B \in \mathcal{B}$, $C(B) \in B_{\mathcal{R}}$. In addition, if $z \in B_{\mathcal{R}}$ and $z \neq C(B)$ then for some issue $\overset{\circ}{B}$, $z = C(\overset{\circ}{B})$ and$\left(B, \overset{\circ}{B}\right)$ is a pair of nested issues at variance. It follows that $(C(B), z)$ is a binary chain. Hence, $C(B)\ \bar{R}'\ z$. It follows that $C(B)$ $\bar{R}'-$optimizes $B_{\mathcal{R}}$. So, $C(B)$ $\bar{R}-$optimizes $B_{\mathcal{R}}$.

The proof of proposition 3 is now as follows: By proposition 2, if x is directly preferred to y, then it must be the case that for any $\hat{R} \in \mathcal{J}^C$, $x\ \hat{R}\ y$. Since $\hat{R}$ is complete and transitive, if x is indirectly preferred to z, then it still must to be the case that $x\ \hat{R}\ z$. This shows part 1 of proposition 3. For part 2 let $(\bar{R}', \mathcal{R})$ be constructed as in the proof of proposition 4. By construction, if x is not (directly or indirectly) revealed as preferred to y then it is *not* the case that $x\ \bar{R}'\ y$. Now consider the binary relation $\bar{R}^{//}$ that is the same as $\bar{R}'$ except that $y\ \bar{R}^{//}\ x$. Given that $\bar{R}'$ is transitive and asymmetric, it follows that $\bar{R}^{//}$ is acyclic and asymmetric.

By topological ordering, $\bar{R}^{//}$ can be extended to an order $\bar{R}$. By construction $\bar{R}$ extends $\bar{R}'$ and $y \; \bar{R} \; x$. By the argument in the proof proposition 4, any order that extends $\bar{R}'$ can underlie C. ■

10.4.5. Proof of Proposition 6

Assume that $(\hat{R}, \mathcal{R})$ underlie C and is not dominated. Then, for every pair $\{x, y\} \subset A$, $\{x, y\}_{\mathcal{R}} = \{x, y\}$. To see this assume, by contradiction, that for some pair of alternatives $\{x, y\}$, $\{x, y\}_{\mathcal{R}} = \{x\}$. Let R' be the rationale such that $y \; R' \; x$ (this is the only relation in R') and let $\mathcal{R}'$ be the set of rationales $\mathcal{R} \bigcup \{R'\}$. By definition, $B_{\mathcal{R}} = B_{\mathcal{R}'}$ for all issues $B \in \mathcal{B}$, $B \neq \{x, y\}$ (because the added rationale only applies to the issue $B = \{x, y\}$). Moreover, $\{x, y\}_{\mathcal{R}'} = \{x, y\}$. This follows because $x \; R \; y$ for some $R \in \mathcal{R}$ (because $x \in \{x, y\}_{\mathcal{R}}$) and $y \in \{x, y\}_{\mathcal{R}'}$ (because $y \; R' \; x$). We now show that $(\tilde{R}^C, \mathcal{R}')$ underlies C.

Let $B \neq \{x, y\}$ be an issue. Let $z \in B_{\mathcal{R}'}$, $z \neq C(B)$. It follows that for some $R \in \mathcal{R}$, $z \; R \; b$ for every $b \in B$ (because $z \in B_{\mathcal{R}}$). In addition, for some $\dot{R} \in \mathcal{R}$, $C(B) \; \dot{R} \; b$ for every $b \in B$. So, $\{C(B), z\}_{\mathcal{R}} = \{C(B), z\}$. It also follows that $C(B)$ $\hat{R} \; z$ (because $z \in B_{\mathcal{R}}$ and $(\hat{R}, \mathcal{R})$ underlie C). Hence, $C\left(\{C(B), z\}\right) = C(B)$ (because $(\hat{R}, \mathcal{R})$ underlie C). By definition, $C(B) \; \tilde{R}^C \; z$. Moreover, $C(\{x, y\}) = x$ (because $C(\{x, y\}) \; \hat{R} \; y$ and $\{x, y\}_{\mathcal{R}'} = \{x, y\}$). Hence, $(\tilde{R}^C, \mathcal{R}')$ underlies C and $(\hat{R}, \mathcal{R})$ is dominated by $(\tilde{R}^C, \mathcal{R}')$. Thus, for every pair of alternatives $\{x, y\}$, $\{x, y\}_{\mathcal{R}} = \{x, y\}$. Given that $(\hat{R}, \mathcal{R})$ underlie C it now follows that $\hat{R} = \tilde{R}^C$. ■

10.4.6. Proof of Proposition 7

Let $\left(\bar{R}, \mathcal{R}\right)$ underlie C. It follows that for any issue $B \in \mathcal{B}$, $C\left(B\right) \in B_{\mathcal{R}}$, and for every $z \in B_{\mathcal{R}}$, $z \neq C\left(B\right)$, $C\left(B\right) \; \bar{R} \; z$. Furthermore, $C\left(B\right) \in B_{\mathcal{R}}$ and $z \in B_{\mathcal{R}}$ imply that there must exist two rationales R_1 and R_2 s.t. $C\left(B\right)$ R_1-optimizes B, and z R_2-optimizes B. Hence, $C\left(B\right) \; R_1 \; z$, and $z \; R_2 \; C\left(B\right)$. Therefore, $\left\{C\left(B\right), z\right\}_{\mathcal{R}} = \left\{C\left(B\right), z\right\}$. Since $C\left(B\right) \; \bar{R} \; z$, it must be the case that $C\left(\left\{C\left(B\right), z\right\}\right) = C\left(B\right)$. Thus, $C\left(B\right) \; \tilde{R}^C \; z$. So, $C\left(B\right)$ $\tilde{R}^C$-optimizes $B_{\mathcal{R}}$ and $\left(\tilde{R}^C, \mathcal{R}\right)$ underlie C.■

10.4.7. Proof of Proposition 8

Proof: Assume, by contradiction, that C is a warm-glow choice function and it does not satisfy WWARP. Then, there exists alternatives $x \neq y$, and issues B_1

and B_2 such that

$$\{x,y\} \subseteq B_1 \subseteq B_2,\ C(\{x,y\}) = C(B_2) = x \text{ and } C(B_1) = y.$$

Let's say that an issue an issue B is resolved by order R if $C(B) = R(B)$.

Step 1. If $(B,\ B^*)$ is at variance then B^* is resolved by R_s (and not by R_m) and B is resolved by R_m (and not by R_s).

Proof: Every issue is resolved by either R_m or R_s. If $(B,\ B^*)$ is at variance then the two issues cannot be resolved by the same rationale. It follows that neither B nor B^* are resolved by both rationales. Consider the case in which B^* is resolved by R_m and B is resolved by R_s. If B^* is resolved by R_m then $R_m(B) = R_m(B^*)$ (because $B \subseteq B^*$ and $R_m(B^*) = C(B^*) \in B$) and $R_m(B^*)\ R_s\ l(R_s(B^*))$ (otherwise B^* is resolved by R_s). Moreover, $R_s(B^*)\ R_s\ R_s(B)$ (because $B \subseteq B^*$). By 10.1, $l(R_s(B^*))\ R_s\ l(R_s(B))$. So, $R_m(B)\ R_s\ l(R_s(B))$. If B is resolved by R_s then $l(R_s(B))\ R_s\ R_m(B)$ (otherwise B^* is resolved by R_m). A contradiction.

The pair of issues $(\{x,y\}, B_1)$ is at variance. By step 1, B_1 is not resolved by R_m. The pair of issues (B_1, B_2) is at variance. By step 1, B_1 is resolved by R_m. A contradiction.■

10.4.8. Proof of Proposition 9

Proof: Assume, by contradiction, that WWARP does not hold. Then there exist $x, y \in B \subset B'$, $x \neq y$, s.t. $C(\{x,y\}) = C(B') = x$, and $C(B) = y$. Let

$$\mathcal{M} = \{M \subseteq B \text{ s.t. } \{x,y\} \subseteq M \text{ and } C(M) = x\}.$$

$\mathcal{M}$ is non-empty as $\{x,y\} \in \mathcal{M}$.

Step 1. There exists $\bar{M} \in \mathcal{M}$ s.t. $\forall T : \bar{M} \subset T \subseteq B$, $C(T) \neq x$.

Let us define the following sequence. For $k = 1$ let $M_1 = \{x,y\} \in \mathcal{M}$, and $p_1 = |M_1| = 2$. For each k, if there exists T s.t. $M_k \subset T \subseteq B$, $C(T) = x$, then we define $M_{k+1} = T \in \mathcal{M}$, and $p_{k+1} = |M_{k+1}| > p_k$. $\{p_k\}$ is a strictly increasing sequence bounded above by the number of elements in B. So, for some $\bar{k}$ there is no T s.t. $M_{\bar{k}} \subset T \subseteq B$ and $C(T) = x$. We define $\bar{M} = M_{\bar{k}}$.

Step 2. $\forall b \in B \backslash \bar{M} : C\left(\bar{M} \cup \{b\}\right) \notin \bar{M}$, and, hence, $C\left(\bar{M} \cup \{b\}\right) = b$.

Suppose not, then there exists $b \in B \backslash \bar{M}$ s.t. $C\left(\bar{M} \cup \{b\}\right) \in \bar{M}$. By definition of $\bar{M}$, $C\left(\bar{M} \cup \{b\}\right) \neq x$, and, therefore, $C\left(\bar{M} \cup \{b\}\right) \notin \left\{x = C\left(\bar{M}\right), b\right\}$. By the EC Axiom, for all issues S s.t. $b \in S$, $C(S) \neq x$. Since $B' \supset B \supset \{b\}$, it follows that $C(B') \neq x$. Contradiction.

Step 3. $\forall T$ s.t. $\bar{M} \subset T \subseteq B$, $C(T) \notin \bar{M}$.

We prove step 3 by induction on the cardinality of T. Step 2 shows that the statement is true for issues T s.t. $|T| = \left|\bar{M}\right| + 1$. Suppose that the statement is true for all issues such that $|T| < \left|\bar{M}\right| + k$, $k \geq 2$. We show that it is true for all issues s.t. $|T| = \left|\bar{M}\right| + k$. Suppose not. Then there exists T s.t. $\bar{M} \subset T \subseteq B$, $|T| = \left|\bar{M}\right| + k$, and $C(T) \in \bar{M}$. Let $b \in T\backslash\bar{M}$. Since $|T\backslash \{b\}| = \left|\bar{M}\right| + k - 1 \geq \left|\bar{M}\right| + 1$, $T\backslash \{b\} \supset \bar{M}$, and $C(T\backslash \{b\}) \notin \bar{M}$. Therefore, $C(T) = C(T\backslash \{b\} \cup \{b\}) \notin \{C(T\backslash \{b\}), b\}$, and, by the EC Axiom, for all issues S s.t. $b \in S$, $C(S) \neq C(T\backslash \{b\})$. Let $m(b) = C(T\backslash \{b\})$. We have showed that for every $b \in T\backslash\bar{M}$, $m(b) \in T\backslash\bar{M}\backslash \{b\}$, and for every $b \in T\backslash\bar{M}$, for all $S \ni b$, $C(S) \neq m(b)$. Now we build the following sequence of elements in $T\backslash\bar{M}$. Let b_1 be any element in $T\backslash\bar{M}$, $b_2 = m(b_1) \neq b_1$, $b_3 = m(b_2)$, and $b_{k+1} = m(b_k)$. Since there is finite number of elements in $T\backslash\bar{M}$, eventually we must form a cycle s.t. $b_j = b_{\bar{j}}$ for some $1 \leq j < \bar{j}$. Since $b_{k+1} = m(b_k)$, b_{k+1} is never chosen when b_k is available. In particular, in a set $\left\{b_j = b_{\bar{j}}, b_{j+1}, ..., b_{\bar{j}-1}\right\}$ none of the elements can be chosen. Contradiction.

From step 3 it follows that $C(B) \notin \bar{M}$, and, so, $C(B) \neq y$. Contradiction.■

References

[1] Akerlof, G. and W. Dickens (1982) "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, 72 (3), 307-319.

[2] Achen, A. and L. Bartels (2006) "It Feels Like We're Thinking: The Rationalizing Voter and Electoral Democracy," mimeo.

[3] Ambrus, A., and K. Rozen. (2008) "Revealed conflicting preferences". Working paper Harvard University.

[4] Andreoni, J. (1989) "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 97, 1447-1458.

[5] Andreoni, J. (1990) "Impure Altruism and Donations to Public Goods. A Theory of Warm-Glow Giving," *Economic Journal*, 100, 464-477.

[6] Andreoni, J. (2006) "Philanthropy" mimeo Wisconsin University.

[7] Becker, G. (1974) "A Theory of Social Interactions," *Journal of Political Economy*, 82, 1063−1093.

[8] Bénabou, R. and J. Tirole (2002) "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117 (3), 871–915.

[9] Berger, P. and G. Smith (1997) "The Effect of Direct Mail Framing Strategies and Segmentation Variables on University Fundraising Performance," *Journal of Direct Marketing*, 2 (1), 30-43.

[10] Bernheim, D. (1984) "Rationalizable Strategic Behavior," *Econometrica*, 52 (4), 1007−1028 .

[11] Bernheim, D. and A. Rangel (2007) "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review*, *papers and proceedings,* 97, 464−470.

[12] Blow, L., M. Browning, and I. Crawford (2008) "Revealed Preference Analysis of Characteristics Models," *Review of Economic Studies*, 75, 371−389.

[13] Bolton, G. and A. Ockenfels (2000) "ERC A Theory of Equity, Reciprocity and Competition," *American Economic Review,* 90 (1), 166−193.

[14] Carillo, J. and T. Mariotti (2000) "Strategic Ignorance as a Self-Discipline Device," *Review of Economic Studies*, 67 (3), 529−544.

[15] Chambers. C. and T. Hayashi (2008) "Choice and Individual Welfare," mimeo Caltech.

[16] Chen, K. (2008) "Rationalization and Cognitive Dissonance: Do Choices Affect or Reflect Preferences," mimeo.

[17] Cormen, T., C. Leiserson, R. Rivest, and C., Stein. (2001) "Introduction to Algorithms." MIT Press and McGraw-Hill. Second Edition.

[18] Dahl, G. and M. Ransom (1999) "Does Where You Stand Depend on Where You Sit? Tithing Donations and Self-Serving Beliefs," *American Economic Review*, 89 (4), 703−727.

[19] Dekel, E., B. Lipman, and A. Rustichini (2001) "Representing Preferences with a Unique Subjective State Space," *Econometrica* 69 (4) , 891−934.

[20] Doyle, J., D. O'Connor, G. Reynolds, and P. Bottomley (1999) "The Robustness of the Asymmetrically Dominated Effect: Buying Frames, Phantom Alternatives, and In-Store Purchases," *Psychology and Marketing*, 16, 225-243.

[21] Easley, D. and A. Rustichini (1999) "Choice Without Beliefs," *Econometrica* 67 (5), 1157–1184.

[22] Eliaz, K. and E. Ok (2006) "Indifference or indecisiveness? Choice-theoretic Foundations of Incomplete Preferences" *Games and Economic Behavior*, 56, 61–86.

[23] Feddersen, T. (2004) "Rational Choice Theory and the Paradox of Voting." *Journal of Economic Perspectives*, 18 (1), 99-112.

[24] Fehr, E., and K. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 817–68.

[25] Fishburn, P. and I. LaValle (1988) "Context-Dependent Choice with Nonlinear and Nontransitive Preferences," *Econometrica*, 56 (5), 1221-1239.

[26] Fudenberg, D. and D. Levine (2006) "A Dual-Self Model of Impulse Control," *American Economic Review*, 96 (5), 1449-1476.

[27] Gale, D. and Mas-Colell, A. (1975) "An equilibrium Existence Theorem for a General Model without Ordered preferences" *Journal of Mathematical Economics* 2, 9–15.

[28] Green, J., and D. Hojman. (2007) "Choice, Rationality and Welfare Measurement". Harvard University. Working Paper.

[29] Gul, F. and W. Pesendorfer (2001) "Temptation and Self-Control," *Econometrica* 69 (6) , 1403–1435.

[30] Gul, F. and W. Pesendorfer (2004) "Self-Control and the Theory of Consumption," *Econometrica* 72 (1) , 119–158.

[31] Gul, F. and W. Pesendorfer (2005) "The Revealed Preference Theory of Changing Tastes," *Review of Economic Studies* 72 (2) , 429–448.

[32] Gul, F. and W. Pesendorfer (2005) "The Case for Mindless Economics," mimeo.

[33] Hubler, J., J. Payne, and C. Puto (1982) “Adding Asymmetrically Dominated Alternatives; Violation of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9, 90-98.

[34] Iyengar, S., and M. Lepper (2000) “When Choice is Demotivating: Can One Desire Too Much of a Good Thing?” *Journal of Personality and Social Psychology*, 79, 995-1006.

[35] Jones, E. (1908) “Rationalisation in Every-day Life,” *Journal of Abnormal Psychology*, 161-169.

[36] Kalai, G., A. Rubinstein, and R. Spiegler (2002) “Rationalizing Choice Functions by Multiple Rationales,” *Econometrica*, 70 (6), 2481-2488.

[37] Maccheroni, F., M. Marinacci and A. Rustichini (2006) “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, 74 (6), 1447-1498.

[38] Manzini, P. and M. Mariotti (2007a) “Sequentially Rationalizable Choice” *American Economic Review* **97**-5, 1824-1839.

[39] Manzini, P. and M. Mariotti. (2007b). “Boundedly Rational Choice, Cycles, and Menu Effects: Theory and Experimental Evidence.” mimeo, Queen Mary, University of London.

[40] Masatlioglu, Y. and D. Nakajima (2007) “A Theory of Choice by Elimination,” mimeo.

[41] Masatlioglu, Y. and E. Ok (2007) “Status Quo Bias and Reference-Dependent Procedural Decision Making,” mimeo.

[42] Mas-Colell, A. (1974) “An equilibrium Existence Theorem without Complete or Transitive Preferences,” *Journal of Mathematical Economics*, 1, 237–246.

[43] May, K (1954) “Intransitivity, Utility, and the Aggregation of Preference Patterns,” *Econometrica*, 22 (1), 1-13.

[44] Mill J.S. (1860): *On Liberty,* P.F. Collier & Son, London.

[45] Moulin, H. (1985) “Choice Functions over a Finite Set: A Summary,” *Social Choice and Welfare*, 2, 147-160.

[46] Mullainathan, S., and R. Thaler (2000) "Behavioral Economics," mimeo.

[47] Ok, E., P. Ortoleva, and G. Riella (2008) "Rational Choice with Endogenous Reference Points," mimeo.

[48] Pearce D. (1984) "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, (4), 1029-1050.

[49] Rabin, M. (1993) "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83 (4), 1281–1302.

[50] Rabin, M. (1995) "Moral Preferences, Moral Constraints, and Self-Serving Biases," mimeo.

[51] Rabin, M. (1998) "Psychology and Economics," *Journal of Economic Literature*, Vol. 36 (1), 11-46.

[52] Roth, A. (2007) "Repugnance as Constraints on Markets," mimeo Harvard University.

[53] Salant, Y. and A. Rubinstein (2006) "Two Comments on the Principle of Revealed Preferences," mimeo.

[54] Salant, Y. and A. Rubinstein (2006a) "A Model of Choice from Lists," *Theoretical Economics*, 1, 3-17.

[55] Samuelson, P. (1938a) "A Note on the Pure Theory of Consumer's Behavior," *Economica*, 5 (17), 61-71.

[56] Samuelson, P. (1938b) "The Empirical Implications of Utility Analysis," *Econometrica*, 6 (4), 344-356.

[57] Sen, A. (1997) "Maximization and the Act of Choice," *Econometrica*, 65 (4), 745-779.

[58] Simon, H. (1959) "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, 49 (3), 253-283.

[59] Simonson, I. (1989) "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *The Journal of Consumer Research*, 16 (2), 158-174.

[60] Simonson, I. and A. Tversky (1992) “Choice in Context: Tradeoff Contrast and Extremeness Aversion,” *Journal of Marketing Research*, 29, 281-295.

[61] Simonson, I and A. Tversky (1993) “Context-Dependent Preferences,” *Management Science*, 39 (10), 1179-1189.

[62] Shafer, W. and H. Sonnenschein (1975) “Equilibrium in Abstract Economies without Ordered Preferences,” *Journal of Mathematical Economics*, 2, 345-348.

[63] Shafir, E., I. Simonson, and A. Tversky (1993) “Reason-based Choice,” *Cognition* 49, 11-36.

[64] Sprumont, Y. (2000) “On the Testable Implications of Collective Choice Theories,” *Journal of Economic Theory*, 93, 205-232.

[65] Spiegler, R. (2002) “Equilibrium in Justifiable Strategies: A Model of Reason-Based Choice in Extensive-Form Games” (2002), *Review of Economic Studies* 69, 691-706.

[66] Spiegler, R. (2004) “Simplicity of Beliefs and Delay Tactics in a Concession Game,” *Games and Economic Behavior* 47 (1), 200-220.

[67] Snyder, M., R. Kleck, A. Strenta, and S. Mentzer (1979) “Avoidance of the Handicapped: An Attributional Ambiguity Analysis,” *Journal of Personality and Social Psychology*, 37 (12), 2297-2306.

[68] Thaler R. and C. Sunstein (2003): “Libertarian Paternalism”, *American Economic Review* 93: 175-179

[69] Tversky, A. and E. Shafir (1992) “Choice Under Conflict: The Dynamics of Deferred Decision,” *Psychological Science*, 3 (6), 358-361.

[70] von Hippel, W., J. Lakin, and R. Shakarchi (2005) “Individual Differences in Motivated Social Cognition: The Case of Self-Serving Information Processing,” *Personality and Social Psychology Bulletin*, 31 (10), 1347-1357.