Sandroni, Alvaro; Cherepavov, Vadim; Feddersen, Timothy

**Article**

# Rationalization

Theoretical Economics

**Provided in Cooperation with:**
The Econometric Society

*Suggested Citation:* Sandroni, Alvaro; Cherepavov, Vadim; Feddersen, Timothy (2013) : Rationalization, Theoretical Economics, ISSN 1555-7561, The Econometric Society, New Haven, CT, Vol. 8, Iss. 3, pp. 775-800, https://doi.org/10.3982/TE970

This Version is available at:
https://hdl.handle.net/10419/150208

Theoretical Economics 8 (2013), 775–800 1555-7561/20130775

# Rationalization

VADIM CHEREPANOV
Department of Economics, University of Pennsylvania

TIMOTHY FEDDERSEN
MEDS Department, Kellogg School of Management, Northwestern University

ALVARO SANDRONI
MEDS Department, Kellogg School of Management, Northwestern University

In 1908, the Welsh neurologist and psychoanalyst Ernest Jones described human beings as *rationalizers* whose behavior is governed by "the necessity of providing an explanation." We construct a formal and testable model of rationalization in which a decision maker selects her preferred alternative from among those that she can rationalize.

KEYWORDS. Rationalization, revealed preferences.

JEL CLASSIFICATION. D01.

## 1. INTRODUCTION

In 1908, the Welsh neurologist and psychoanalyst Ernest Jones wrote a paper titled "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 Sigmund Freud with the critical insight "that a number of mental processes owe their origin to causes unknown to and unsuspected by the individual" he writes that *rationalization* occurs because people feel "*a necessity to provide an explanation*" (Jones 1908).

The idea of rationalization has become so well accepted that pundits write about it in the popular press. Psychologists emphasize the facility with which people create implausible explanations for their behavior. However, the phenomena of rationalization can only influence choice if the inability to rationalize constrains behavior.

Vadim Cherepanov: vadimch@sas.upenn.edu
Timothy Feddersen: tfed@kellogg.northwestern.edu
Alvaro Sandroni: sandroni@kellogg.northwestern.edu

We thank the co-editor, Bart Lipman, and two anonymous referees for useful comments and suggestions. We also thank Eddie Dekel, Jennifer Jordan, Paola Manzini, Marco Mariotti, Herve Moulin, Efe Ok, Nicola Persico, Charlie Plott, Scott Presti, Yuval Salant, Yves Sprumont, and Marciano Siniscalchi for useful comments. Sandroni gratefully acknowledges financial support from the National Science Foundation Grant 0922404. All errors are ours.


DOI: 10.3982/TE970

We seek to better understand the logic of rationalization by developing a formal theory. Our main premise is that agents choose according to their preferences, but face potentially unobservable *psychological constraints*. For example, a manager may have the opportunity and incentive to commit fraud, but absent a convincing rationale that legitimizes fraud, may be psychologically constrained not to commit fraud and does not do so.

We model a decision maker (Dee) who has preferences over alternatives and a set of *rationales* (modeled as binary relations). Dee chooses the alternative she prefers among the feasible options she can *rationalize* i.e., those that are optimal according to at least one of her rationales. To rationalize a choice is, therefore, to find a subjectively appealing rationale that justifies that choice. So rationalization is a constrained optimization process with (possibly unobservable) constraints.

Consider the following example: given the choice between work ($w$) and a movie ($m$), Dee chooses the movie. However, when Dee has a third option of visiting a relative in the hospital ($h$) she stays at work. Dee's choice of $m$ from $\{w, m\}$ and $w$ from $\{w, m, h\}$ violates the *Weak Axiom of Revealed Preferences* (WARP) (see Samuelson 1938). Rationalization theory accommodates her behavior. Suppose that Dee prefers the movie to work and prefers work to visiting the hospital. She has two rationales available to justify her choices. Under rationale 1, Dee's work is pressing and $w$ is ranked above $m$ and $h$. Under rationale 2, work is not pressing and $h$ is ranked above $m$ and $m$ is ranked above $w$. Dee chooses the movie over work because she prefers it and rationalizes her choice using rationale 2. However, between all three options, Dee chooses work because she cannot rationalize her preferred choice (the movie) but can rationalize her second choice of work using rationale 1.

Some behavioral anomalies can be accommodated by rationalization theory. Indeed, Dee's preferences can be a stable, single order even if observed choices are cyclic. Nevertheless, rationalization theory is testable. A known axiom akin to WARP characterizes the empirical content of rationalization theory.

One goal of this paper is to develop a formal theory that can capture the logic of rationalization and help interpret data. For example, a literature in social psychology investigates responses to stigmatized groups. Snyder et al. (1979) allow subjects to choose whether to watch a movie alone or with someone in a wheelchair. In one treatment, subjects disproportionately choose to watch a movie with a person in a wheelchair rather than watching the same movie alone. In a second treatment, when the movies are different, subjects disproportionately choose to watch a movie alone rather than with the handicapped person. The experiment was designed to rule out actual preferences between the movies as an explanation for behavior. In the handicapped avoidance claim, subjects want to avoid the handicapped, but choose not to. In the first treatment, subjects are psychologically constrained to watch the movie with the handicapped person because to do otherwise would require subjects to reveal (perhaps only to themselves) handicapped aversion.

Rationalization theory captures the behavior in Snyder's study in a way that is consistent with the handicapped avoidance claim: Dee prefers to see the movie alone, but can-

not rationalize doing so when the movies are the same. In the first treatment, Dee acts against her preferences because of a psychological constraint, while in the second treatment, the constraint is relaxed. She can rationalize watching the second movie alone by telling herself that she prefers that movie. The implication is that legitimizing behavior (such as discrimination) may remove psychological constraints that result in changes in conduct.

However, the observed behavior is also consistent with preferences and psychological constraints that do not require handicapped aversion. The handicapped avoidance claim follows from cyclic choices, ordered preferences, and the assumption that subjects can rationalize seeing a movie with the handicapped person instead of watching a different movie alone. This last assumption is a *permissibility assumption*.

A permissibility assumption stipulates that Dee has a rationale that allows her to choose a given option in a given set of alternatives. The use of permissibility assumptions is akin to the use of nonchoice data and typically is avoided by decision theorists. However, as our handicapped aversion example shows, these assumptions may underlie intuitive deductions about preferences.

We show inferences about preferences and constraints that can be made with and without permissibility assumptions. A preference order is *identifiable* from observed choices if there are permissibility assumptions that pin it down uniquely. We show that a preference order is identifiable if and only if it accommodates Dee's choices while imposing minimal psychological constraints on what she might choose. So if a preference order requires more constraints than needed to accommodate choices, then it cannot be shown to be Dee's order under any assumptions over her constraints.

We define the *minimum constraint theory of rationalization* as the set of rationalization models that does not require more constraints than needed to accommodate choices. If behavior is not anomalous, the minimum constraint theory of rationalization reveals the same preference order as standard economic theory. However, the minimum constraint rationalization theory also reveals preferences and constraints in settings in which choice is anomalous. So the minimum constraint theory of rationalization extends standard theory. It uniquely reveals a standard preference order when behavior is not anomalous and allows for precise inferences of preferences in cases where standard economics makes contradictory inferences about preferences.

The organization of the paper is as follows. Section 2 provides a brief literature review. Section 3 formalizes the idea of rationalization. Section 4 shows results on empirical content and revealed preferences of rationalization theory, and some of the implications of these results for empirical work. Section 5 introduces the minimum constraint theory of rationalization. Section 6 shows a detailed comparison of our model with alternative theories of choice constraints and some directions for future work. Section 7 provides a conclusion. Proofs are given in the Appendix.

## 2. Related literature

A growing literature focuses on conflicting motivations. See, among many contributions, Ambrus and Rozen (2008), Chambers and Hayashi (2012), de Clippel and Eliaz

(2012), Dietrich and List (2010), Eliaz and Ok (2006), Fudenberg and Levine (2006), Gul and Pesendorfer (2005), Kalai et al. (2002), Heller (2012), Lehrer and Teper (2011), Masatlioglu and Nakajima (2007), Ok et al. (2012), and Rubinstein and Salant (2006a). While these models do not formalize the idea of rationalization, they accommodate behavioral anomalies.

The word "rationalizability" is used in game theory (see Bernheim 1984, Pearce 1984, Sprumont 2000) differently from us. The word "rationalization" is also used differently in cognitive dissonance theory. The basic claim is that people devalue rejected choices and valorize chosen ones (see Chen 2008). In the area of motivated cognition, von Hippel et al. (2005) provides a survey on self-serving biased information processing (see also Akerlof and Dickens 1982, Rabin 1995, Carillo 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.

Well known experimental work provides evidence that is consistent with psychological legitimation and delegitimation of options (see, among many contributions, Gneezy and Rustichini 2000, Mazar and Ariely 2006). Roth (2007) lists potentially beneficial practices that were deemed repugnant and banned. Examples include the human consumption of horse meat (illegal in California), selling pollution permits, and markets for human organs. Emotions such as repugnance may affect preferences and constraints. So we identify both preferences and constraints from choice and, sometimes, nonchoice data as well.

Spiegler (2002, 2004) develops game-theoretic models in which players must justify their chosen actions. The models closest to ours consider constraints on choice beyond feasibility. In Section 7, we show similarities and differences between alternative theories of behavior and ours. To simplify the exposition, we focus this comparative analysis on the work of Manzini and Mariotti (2007, 2012a), Masatlioglu et al. (2012), and Lleras et al. (2010).

## 3. Basic concepts

Let $A$ be a finite set of alternatives. A nonempty 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 $B \in \mathcal{B}$. A decision maker named Dee makes the choices given by $C$.

A *preference* $P$ is an asymmetric binary relation on $A$. A transitive, complete preference is an *order*. We emphasize the case in which Dee's preferences are orders. In a few instances (e.g., for a comparison with the work of Manzini and Mariotti 2007), we make the weaker assumption of asymmetry only. As usual, $x\ P\ y$ denotes that $x$ is $P$-preferred to $y$. Let $\mathcal{P}$ be the set of preferences and let $\mathcal{P}^o \subseteq \mathcal{P}$ be the set of preferences orders.

A *psychological constraint function* is a mapping $\psi : \mathcal{B} \longrightarrow \mathcal{B}$ such that $\varnothing \neq \psi(B) \subseteq B$ for every issue $B \in \mathcal{B}$. An option $x \in \psi(B)$ is *psychologically permissible in issue* $B$. Let $\Psi$ be the set of psychological constraint functions.

A *model of behavior* is a pair $(P, \psi)$ of a preference and a psychological constraint function. A model of behavior $(P, \psi)$ *underlies* a choice function $C$ if, for any issue $B \in \mathcal{B}$, $C(B) \in \psi(B)$ and

$$C(B)\, P\, y \quad \text{for all } y \in \psi(B),\ y \neq C(B).$$

Dee chooses the option she prefers among the psychologically permissible options.

Given a choice function $C$, a *theory of behavior* is a subset $\widehat{\mathcal{P} \times \Psi} \subseteq \mathcal{P} \times \Psi$ of preferences and psychological constraint functions. So a theory of behavior is a collection of models of behavior.[1] A choice function $C$ is *consistent* with a theory of behavior $\widehat{\mathcal{P} \times \Psi}$ if some model of behavior $(P, \psi) \in \widehat{\mathcal{P} \times \Psi}$ underlies $C$.

A binary relation (not necessarily complete, transitive, or asymmetric) $R$ on $A$ is called a *rationale*. A rationale can be intuitively understood as a story (which may differ from Dee's preferences) that states that some options are better than others. Given an issue $B$, an alternative $x \in B$ is *rationalized* by $R$ if and only if $x\, R\, y$ for all $y \in B$, $y \neq x$. So $R$ rationalizes $x$ if it tells Dee that $x$ is the best course of action.

Dee can use any story that she can accept to rationalize an option. Let $\mathcal{R} = \{R_i,\ i = 1, \ldots, n\}$ be the set of Dee's rationales. Given $\mathcal{R}$, an option $x \in B$ is *rationalizable* in $B$ if $x$ is rationalized by a rationale $R_i \in \mathcal{R}$ that Dee accepts. To rationalize an option $x$, Dee requires only that one of her rationales ranks $x$ as the best course of action. Other rationales may regard $x$ as an inferior option. A set of rationales $\mathcal{R}$ defines a psychological constraint function $\psi^{\mathcal{R}}$, where $\psi^{\mathcal{R}}(B)$ is the set of rationalizable options in $B$. Let $\bar{\Psi} \subseteq \Psi$ be the set of all psychological constraint functions $\psi$ such that $\psi = \psi^{\mathcal{R}}$ for some set of rationales $\mathcal{R}$. Let $\mathcal{P} \times \bar{\Psi}$ be the *basic theory of rationalization* and let $\mathcal{P}^o \times \bar{\Psi}$ be the *theory of order rationalization*, i.e., rationalization theory with ordered preferences.

In an alternative definition of rationalization, an option $x \in B$ is said to be rationalized by $R$ if and only if no alternative $y \in B$ is such that $y\, R\, x$. Our results are the same under both definitions of rationalization (see the Appendix for a proof).[2]

## 4. Empirical content of rationalization theory

Weak WARP. A choice function $C$ satisfies the Weak WARP condition if and only if

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

Weak WARP is a relaxation of WARP (see Manzini and Mariotti 2007, 2012a and Ehlers and Sprumont 2008). Given a choice function $C$, choices $C(B)$ and $C(B^*)$ are *anomalous* if

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

[1] The dependence on the choice function is only useful in Section 5, where we introduce the idea of minimum constraint to select models according to the evidence. Thus, with some abuse of notation, we may refer to a theory as just a subset $\widehat{\mathcal{P} \times \Psi}$ of $\mathcal{P} \times \Psi$.

[2] In another theory (see the Appendix), Dee selects among options for which she finds reasons to eliminate the alternatives.

It is convenient to define $x$ *Rev* $y$ if there exist anomalous choices $C(B)$ and $C(B^*)$, $B \subseteq B^*$, such that $x = C(B)$ and $y = C(B^*)$.

No Binary Chain Cycles. A choice function $C$ satisfies No Binary Chain Cycles if and only if there is no chain of alternatives $\{x_0, \ldots, x_n\}$ such that $x_0 = x_n$ and $x_i$ *Rev* $x_{i+1}$, $i = 0, \ldots, n-1$.

Proposition 1. *A choice function $C$ is consistent with basic rationalization theory if and only if it satisfies Weak WARP. A choice function $C$ is consistent with order rationalization theory if and only if it satisfies No Binary Chain Cycles.*

So order rationalization theory can accommodate all behavioral anomalies that satisfy No Binary Chain Cycles and, therefore, is consistent with any observed choices over three alternatives. In particular, Dee's choices can be cyclic even if her preferences are ordered. Finally, rationalization theory is testable in the sense that we can identify behavior that is inconsistent with the theory.

### 4.1 *Intuition behind Proposition 1 and related results*

If constraints are unobservable and unrestricted, then any choice of $x$ in issue $B$ can be accommodated by a model in which Dee's choice is dictated by her constraints, i.e., $\psi(B) = \{x\}$. A testable model of behavior must put some structure on psychological constraints. Dee's rationales need not be transitive, complete, or asymmetric and yet the need to rationalize imposes structure on psychological constraints.

Structure on psychological constraints imposed by rationalization. A psychological constraint function $\psi$ belongs to $\bar{\Psi}$ if and only if it satisfies

$$\text{if } B \subseteq B^* \quad \text{then } \psi\big(B^*\big) \cap B \subseteq \psi(B). \tag{1}$$

So, (1) is the structure on psychological constraints required by rationalization. The intuition behind this result is that if Dee can rationalize an option $x$ in $B^*$, then she can also rationalize it in a subset $B$ of those alternatives. That is, as the set of alternatives grows, it becomes harder to justify $x$ as the right course of action. For example, Dee may be able to rationalize killing someone if her only alternative option is to die, but not if she can avoid any death by taking another action.

The intuition in Proposition 1 is as follows: Assume that $x$ *Rev* $y$. A choice of $y$ in $B^*$ shows that $y$ is permissible in $B^*$ and, therefore, it is also permissible in $B \subseteq B^*$. If $y$ is permissible in $B$, then the choice of $x$ in $B$ reveals a preference for $x$ over $y$. The role of Weak WARP is to rule out contradictory inferences of preferences of the form $x$ *Rev* $y$ and $y$ *Rev* $x$. The role of No Binary Chain Cycles is to avoid cyclic revealed preferences such as $x$ *Rev* $y$, $y$ *Rev* $z$, and $z$ *Rev* $x$.

There is no loss of generality in restricting rationales to be transitive and asymmetric. If $\psi$ satisfies (1), then there is a set $\mathcal{R}$ of rationales such that $\psi = \psi^{\mathcal{R}}$, and each rationale

in $\mathcal{R}$ is transitive and asymmetric (see the Appendix). However, if rationales must be orders, then the set of psychological constraint functions is strictly contained in $\Psi^{\mathcal{R}}$.

An example (from Sen 1997) illustrates the limits of (1). Assume that Dee must choose a chair and finds it psychologically difficult to pick the most comfortable. While such a psychological constraint is understandable, it is ruled out by (1) because Dee can justify the second most comfortable chair when the most comfortable chair is available, but not otherwise. We elaborate on this point in Section 6 and conclude this section with formal results on revealed preferences.

DEFINITION 1. Given choice function $C$ and theory $\widehat{\mathcal{P} \times \Psi}$, $x$ is revealed to be preferred to $y$ if $C$ is consistent with $\widehat{\mathcal{P} \times \Psi}$ and $x\ P\ y$ in every model of behavior $(P, \psi) \in \widehat{\mathcal{P} \times \Psi}$ that underlies $C$.

So $x$ is revealed to be preferred to $y$ if $x$ ranks higher than $y$ in every model of behavior (within a given theory) that underlies Dee's choices.

PROPOSITION 2. *Let $C$ be a choice function consistent with basic rationalization theory $\mathcal{P} \times \bar{\Psi}$. Option $x$ is revealed to be preferred to $y \neq x$ if and only if $x$ Rev $y$. Under order rationalization theory, $x$ is revealed to be preferred to $y \neq x$ if and only if there is a chain of alternatives $\{x_0, \ldots, x_n\}$ such that $x_0 = x$, $y = x_n$, and $x_i$ Rev $x_{i+1}$, $i = 0, \ldots, n-1$.*

### 4.2 *Avoidance of the handicapped*

As mentioned in the Introduction, Snyder et al. (1979) design an experiment with the alternatives watch movie 1 alone ($x$), watch movie 2 alone ($y$), and watch movie 1 with a person in a wheelchair ($z$). Several subjects choose to watch movie 1 with the handicapped rather than movie 1 alone (i.e., $\bar{C}(x, z) = z$). Many subjects choose to watch movie 2 alone rather than movie 1 with the handicapped (i.e., $\bar{C}(y, z) = y$).

The handicapped avoidance claim is that some subjects prefer to avoid the handicapped (i.e., they prefer $x$ to $z$), but choose $z$ rather than $x$. To put this claim in terms of rationalization theory, subjects cannot rationalize what they prefer (to see the movie alone) when the movies are identical, but they can rationalize watching the movie alone when the movies are different (perhaps by telling themselves that they prefer the movie that they can see alone).

Consider the choice between movie 1 and movie 2 (i.e., between $x$ and $y$). Some choose $x$ and some choose $y$. The behavior of those who choose $y$ can be accommodated without handicapped aversion or psychological constraints. These subjects may prefer $y$ to $z$ to $x$. That is, they saw the movie 1 alone as opposed to movie 2 with the person in a wheelchair because they prefer movie 1 to movie 2. Alternatively, consider those who choose $x$ over $y$ (i.e., $\bar{C}(x, y) = x$). Now the observed choice behavior is cyclic: $x$ chosen over $y$, $y$ chosen over $z$, and $z$ chosen over $x$. This cycle is still not sufficient to show the handicapped aversion claim. To see this, consider these choices and assume that Dee chooses to see movie 2 alone when all three options are available (i.e.,

$\bar{C}(x, y, z) = y$).[3] By Proposition 2, it follows that Dee prefers movie 1 to movie 2 (i.e., Dee prefers $x$ to $y$). Thus, the choice to watch movie 2 alone as opposed to movie 1 with the person in the wheelchair is *not* due to a preference for movie 2. However, also by Proposition 2, the handicapped avoidance claim (Dee prefers $x$ to $z$) does *not* follow because no matter which option is chosen when all three options $x$, $y$, and $z$ are available, it does *not* follow that $x$ is preferred to $z$ because $x$ was *not* chosen over $z$ in the binary choice. So the handicapped avoidance claim may be one way to accommodate the evidence, but there are other ways to accommodate the same evidence as well. That is, even if subjects prefer movie 2 to movie 1 and choose to watch movie 1 alone over movie 2 with the handicapped person, it does *not* follow that they prefer to avoid the handicapped. We give some intuition for this point below.

The handicapped avoidance claim can be captured by a rationalization model, let us call it the $S$-model, where Dee's preference order is $x$ to $y$ to $z$ and the psychological constraints are $\psi\{x, z\} = \{z\}$, $\psi\{x, y, z\} = \{y, z\}$, and $\psi\{B\} = B$ in all other issues. The $S$-model underlies the choices in $\bar{C}$ and captures the intuition that Dee prefers to avoid the handicapped ($x$ preferred to $z$), but cannot rationalize her preferred choice ($\psi\{x, z\} = \{z\}$) when the movies are the same; she can only rationalize avoiding the handicapped when the movies differ. However, consider another model, the $S'$-model, where the preference order is $z$ to $x$ to $y$; $\psi'\{y, z\} = \psi'\{x, y, z\} = \{y\}$, $\psi'\{B\} = B$ elsewhere. The $S'$-model accommodates the choices in $\bar{C}$ even though it does not involve handicapped aversion. Both the $S'$-model and the $S$-model are based on the same preferences over the movies (movie 1 is better than movie 2). So the $S'$-model is a way to accommodate the data that are not based on handicapped aversion or unusual preferences over the movies. It follows that the handicapped avoidance result cannot yet be obtained.

## 5. The minimum constraint theory of rationalization

Additional assumptions may be made on psychological constraints. For example, if Dee belongs to a religious organization that valorizes helping the needy, then it may be justified to assume that helping the needy is psychologically permissible for Dee. Formally, *permissibility assumptions* are a set $\mathcal{A} = \{(y_i, B_i); y_i \in B_i \ i = 1, \ldots, n\}$ of $n$ issues $B_i \in \mathcal{B}$ and alternatives $y_i \in B_i$, implying that $y_i$ is psychologically permissible in $B_i$. Let $\Psi^{\mathcal{A}} \subseteq \bar{\Psi}$ be the set of all psychological constraint functions $\psi$ that satisfy (1) and such that $y_i \in \psi(B_i)$, $i = 1, \ldots, n$. Let $\mathcal{P}^o \times \Psi^{\mathcal{A}}$ be the *theory of order $\mathcal{A}$-rationalization*. The empirical content of this theory is shown in the Appendix.

Definition 2. Given a choice function $C$, theory $\widehat{\mathcal{P} \times \Psi}$ identifies a preference order $\dot{P}$ if some model in $\widehat{\mathcal{P} \times \Psi}$ underlies $C$ and $P = \dot{P}$ for any model $(P, \psi) \in \widehat{\mathcal{P} \times \Psi}$ that underlies $C$. A preference order $\dot{P}$ is identifiable if there exist permissibility assumptions $\mathcal{A}$ such that the theory of order $\mathcal{A}$-rationalization identifies $\dot{P}$.

[3] We do not follow the exact protocol in Snyder et al. (e.g., they do not show a choice over three alternatives).

So a preference is identifiable if for *some* theory of behavior (based on preference orders and permissibility assumptions), this preference is the only one that can accommodate the observed choices.

Let $(P, \psi)$ and $(P', \psi')$ be two models that underlie a choice function $C$. The model $(P, \psi)$ is *dominated* by $(P', \psi')$ if $P'$ is an order and $\psi(B) \subseteq \psi'(B)$ for all issues $B \in \mathcal{B}$, with strict inclusion for some issue. So a dominated model $(P, \psi)$ uses more constraints than necessary to accommodate the choices. Given a choice function $C$, the *minimum constraint theory of rationalization* consists of all models $(P, \psi) \in \mathcal{P}^o \times \bar{\Psi}$ that underlie $C$ and are not dominated by an alternative model $(P', \psi') \in \mathcal{P}^o \times \bar{\Psi}$. In the appendix (Proposition A.2), we show that if a choice function is consistent with order rationalization theory, then it is also consistent with minimum constraint theory of rationalization. So this theory does not produce an empty set of models.

THEOREM 1. *Given a choice function $C$, a preference order $P$ is identifiable if and only if there is a psychological constraint function $\psi$ such that the model $(P, \psi)$ belongs to the minimum constraint theory of rationalization.*

Theorem 1 shows that a preference is identifiable from choice and permissibility assumptions if and only if it belongs to a model that accommodates choices without using more constraints than needed. This result supports the idea of selecting minimum constraint models: they are the only ones based on identifiable preferences.

The intuition in Theorem 1 is that if a model $(P, \psi)$ is dominated, then there is an alternative model $(P', \psi')$ with fewer constraints that also accommodates the observed choices. So any permissibility assumption satisfied by $\psi$ is also satisfied by $\psi'$. Thus, $(P, \psi)$ cannot be the only way to accommodate the choices.

Note that $(y, B) \notin \mathcal{A}$ means that $y$ is not assumed to be permissible in $B$. It does not mean that $y$ is assumed to be impermissible in $B$. An *impermissibility assumption* stipulates that Dee is psychologically constrained from choosing some alternatives in specific issues. That is, an impermissibility assumption $\mathcal{T}$ is a set of issues $B_j$ and options $y_j \in B_j$ such that $y_j \notin \psi(B_i)$. Impermissibility assumptions do not help identify preferences. Consider any $(y_j, B_j) \in \mathcal{T}$ and any choice function $C$. If $y_j = C(B_j)$, then Dee's choice contradicts the assumption that $y_j$ is impermissible in $B_j$. Alternatively, if $y_j \neq C(B_j)$ for every $j$, then it is straightforward to show that given any model $(P, \psi)$ that underlies $C$, there exists an alternative model $(P, \psi')$ that also underlies $C$ such that $\psi'$ satisfies $\mathcal{T}$ (i.e., $\psi'(B)$ comprise all options in $\psi(B)$ minus those that are assumed by $\mathcal{T}$ to be impermissible).

If the choice function $C$ satisfies WARP, then the minimum constraint theory of rationalization identifies the same preference order as the standard theory of choice. Thus, the minimum constraint theory of rationalization does not modify standard economics, but it may reveal preferences from anomalous behavioral patterns. Let us say that a choice function is *acyclic*if the binary choices do not form a cycle.

PROPOSITION 3. *Let $C$ be an acyclic choice function that is consistent with rationalization theory. If $C(\{x, y\}) = x$, $x \neq y$ (i.e., $x$ is chosen over $y$ in a binary choice), then $x$ is revealed to be preferred to $y$ by minimum constraint rationalization theory.*

Proposition 3 shows that when a choice function is acyclic, then the only surviving preference relation is the order defined by the binary choices. Thus, the identification of preferences is broadened to include some anomalous behavior.

PROPOSITION 4. *Let* $(P,\psi) \in \mathcal{P}^o \times \bar{\Psi}$ *and* $(P,\psi') \in \mathcal{P}^o \times \bar{\Psi}$ *be two models that underlie choices* $C$ *and are not dominated by any alternative model in* $\mathcal{P}^o \times \bar{\Psi}$. *Then* $\psi = \psi'$.

By Proposition 4, if Dee's preferences are revealed under the minimum constraint rationalization theory, then Dee's constraints are also identified. Thus, Propositions 3 and 4 reveal preferences and constraints when observed binary choices are acyclic.

### 5.1 *Difficult choice anomaly*

Consider the choice function $C(e_1,e_2) = e_1$, $C(e_1,n) = e_1$, $C(e_2,n) = e_2$, and $C(e_1,e_2,n) = n$. This pattern is anomalous because $n$ is rejected over $e_1$ and also over $e_2$ separately, but $n$ is chosen over both $e_1$ and $e_2$ when they are simultaneously available. From this pattern of choice alone, we can determine preferences and constraints. By Proposition 3, minimum constraint rationalization theory reveals that $e_1\ P\ e_2\ P\ n$.[4] Proposition 4 reveals that the only binding psychological constraint occurs in the issue $\{e_1,e_2,n\}$ in which neither $e_1$ nor $e_2$ is psychologically permissible.

The choices above are consistent with an anecdote about Thomas Schelling (as told by Shafir et al. 1993), who, on one occasion, had decided to buy an encyclopedia. Upon arriving at the bookstore, had only one encyclopedia been available ($e_1$ or $e_2$), he would have bought it. However, he was presented with two encyclopedias and bought none ($n$). In our interpretation, Schelling found it hard to explain to himself why one encyclopedia is better than the other.

The pattern of behavior above is an acyclic behavioral anomaly often called a *difficult choice*. Tversky and Shafir (1992) and Tversky and Simonson (1993), among others, noted the difficult choice anomaly in several experiments. In a field experiment, Iyengar and Lepper (2000) observe that the fraction of customers who buy a gourmet jam is significantly larger when customers are presented with a limited selection than with an extensive selection.

### 5.2 *Cycles*

Consider the three-alternative cycle: $x$ over $y$, $y$ over $z$, and $z$ over $x$ (let us say $y$ is chosen when all three alternatives are available). This is the behavioral pattern $\bar{C}$ in the handicapped aversion example. Both the $S$-model and the $S'$-model are undominated and underlie $\bar{C}$. As shown by Theorem 1, each undominated model can be differentiated by permissibility assumptions. Thus, the handicapped avoidance claim may depend on judgement over these permissibility assumptions.

Consider the assumption that $x$ is rationalizable in $(x,z)$ (i.e., $x \in \psi(x,z)$). That is, Dee can rationalize watching a movie alone when the alternative is to see the same

[4]The inference that $e_1\ P\ n$ and $e_2\ P\ n$ follows directly from Proposition 2 and does not require minimum constraint theory of rationalization.

movie with a handicapped person. Then, from $z$ chosen over $x$, it follows that Dee prefers $z$ to $x$. By Proposition 2, Dee prefers $x$ to $y$. So Dee's revealed preference order is model $S'$: she prefers $z$ to $x$ to $y$; she does not prefer to avoid the handicapped. Now consider the assumption that $z$ is permissible in $(y,z)$ (i.e., $z \in \psi(y,z)$). That is, Dee can rationalize watching a movie with the handicapped person when the alternative is to see another movie alone. Then, from $y$ chosen over $z$, it follows that Dee prefers $y$ to $z$. By Proposition 2, Dee prefers $x$ to $y$. So, Dee's revealed preference order is model $S$: she prefers $x$ to $y$ to $z$; she prefers to avoid the handicapped. So under the choice function $\bar{C}$, the handicapped aversion claim follows from the permissibility assumption $z \in \psi(y,z)$. Finally, if $C(x,y) = x$ and $C(x,y,z) = y$, then the handicapped avoidance claim also follows under the assumption that all rationales must be ordered.[5]

## 6. Differentiating theories

Rationalization theory belongs to a family of theories where Dee optimizes, or maximizes, her preferences over a subset $\psi(B)$ of her feasible options $B$ (see, among many contributions, Eliaz et al. 2011, Eliaz and Spiegler 2011, and Masatlioglu and Nakajima 2007). To ease comparisons, we focus on the work of Manzini and Mariotti (2007, 2012a), Masatlioglu et al. (2012), and Lleras et al. (2010). We now present the constraint functions in these papers.[6]

The constraint functions in Manzini and Mariotti (2007) are mappings $\psi^{\mathrm{SL}}: \mathcal{B} \longrightarrow \mathcal{B}$ such that for some preference $P_1$,

$$\psi^{\mathrm{SL}}(B) = \{x \in B \mid \nexists y \in B \text{ for which } y \, P_1 \, x\}. \tag{2}$$

So a constraint function in Manzini and Mariotti (2007) is such that all permitted options are maximal according to some asymmetric binary relation $P_1$. Dee decides in two stages. In the first stage, she eliminates some options (using $P_1$). In the second stage, she chooses a maximal option (using a preference relation $P_2$). For example, in the process of buying a car, Dee may first decide to buy an American car. So $P_1$ ranks Americans cars above foreign cars. In the second stage, Dee selects an American car that she likes best according to her preference relation $P_2$. This two-stage decision process is called a *rational short list*.

In the Appendix, we show that if a constraint function satisfies (2), then it also satisfies (1), but the converse does not hold. Some constraint functions satisfy (1), but not (2). Thus, the constraint functions in a rational short list are a special case of the constraint functions in rationalization theory. So the choice functions accommodated by a rational short list can also be accommodated by rationalization theory.

[5]Assume, to the contrary, that Dee prefers $z$ to $x$. By Proposition 2, Dee must prefer $z$ to $x$ to $y$. From $C(x,y,z) = y$, it follows that all ordered rationales must rank $y$ highest. So $x$ cannot be rationalized if $y$ is feasible. This contradicts $C(x,y) = x$.

[6]In this section, we refer to psychological constraint functions as constraint functions because now we consider different theories and so we require a broader interpretation of these constraints. The sets $\psi(B)$ are sometimes called consideration sets.

Building on their original work, Manzini and Mariotti (2012a) develop a more flexible theory of categorization, called *categorize then choose*, that is characterized by Weak WARP. Hence, the empirical scope of the categorize-then-choose theory coincides with the empirical scope of basic rationalization theory. Under the assumption that Dee's preferences are orders, Manzini and Mariotti (2012a) show that the empirical scope of categorize then choose is strictly subsumed by the empirical scope of order rationalization theory. So under the assumption of ordered preferences, rationalization theory can be empirically distinguished from the categorize-then-choose theory based on choice data alone.

The constraint functions in Masatlioglu et al. (2012), called *attention filters*, are mappings $\psi^{\mathrm{AF}}:\mathcal{B}\longrightarrow\mathcal{B}$ such that for any issue $B$, $\psi^{\mathrm{AF}}(B)\subset B$ and

$$\psi^{\mathrm{AF}}(B)=\psi^{\mathrm{AF}}(B\setminus x)\quad\text{whenever } x\notin\psi^{\mathrm{AF}}(B). \tag{3}$$

In the work of Masatlioglu et al. (2012), Dee may not pay attention to all her available options in $B$ and, instead, focus on a subset $\psi^{\mathrm{AF}}(B)$. Equation (3) ensures that if Dee does not pay attention to option $x$ (i.e., $x\notin\psi^{\mathrm{AF}}(B)$), then removing option $x$ does not alter the options she pays attention to. Dee optimizes over $\psi^{\mathrm{AF}}(B)$ with her preference order. This process is called *choice with limited attention*.

There are constraint functions that satisfy (1) but not (3) and, conversely, constraint functions that satisfy (3) but not (1) (see Lleras et al. 2012, Example 2). Thus, attention filters may not satisfy the structure imposed by rationalization and, conversely, rationalization constraint functions may not be attention filters. In addition, there are choice functions with limited attention that do not satisfy Weak WARP and, conversely, there are order rationalization choice functions that are not choice functions with limited attention (see Examples 2 and 3 in Masatlioglu et al. 2012). Thus, the empirical scope of these two theories differ. In addition, in our leading example of hidden discrimination, Dee must prefer to watch movie 1 alone rather than to watch movie 1 with a person in a wheelchair. This inference was shown under order rationalization theory and a suitable permissibility assumption. However, it is straightforward to show (a proof is available on request) that the same preference inference does not follow under the Masatlioglu–Nakajima–Ozbay model of limited attention, even if the same choice function $\bar{C}$ is observed and the same permissibility assumption is made.

The constraint functions in Lleras et al. (2010), called *consideration filters*, are mappings $\psi^{\mathrm{CF}}:\mathcal{B}\longrightarrow\mathcal{B}$ that satisfy $\psi^{\mathrm{AF}}(B)\subset B$ and (1). The intuitive idea in Lleras et al. (2010) is also that Dee may not pay attention to all her available options in $B$ and, instead, focus on a subset $\psi^{\mathrm{CF}}(B)$. It is assumed that if Dee pays attention to an option $x$ in a issue $B^*$, then she also pays attention to an option $x$ in a subset $B\subseteq B^*$. Dee optimizes over $\psi^{\mathrm{CF}}(B)$ with her preference order. This decision process is called a *choice with limited consideration*. Thus, while order rationalization theory and limited consideration theory were designed to conceptualize different intuitive ideas, they have the same formal structure.

Consider the constraint function $\bar{\psi}$ such that $\bar{\psi}(B)$ consists of a single option in $B$ and $B\subseteq B^*$, and $\bar{\psi}(B^*)\in B\Longrightarrow\bar{\psi}(B^*)=\bar{\psi}(B)$. So if the option $\bar{\psi}(B)\in B$ is seen as a

"choice," then WARP is satisfied. Note that $\bar{\psi}$ satisfies (1), (2), and (3). Moreover, if a choice function $C$ is not anomalous, then it can be accommodated by a model $(P, \bar{\psi})$, where $\bar{\psi}(B) = \{C(B)\}$ and $P$ is an arbitrary preference. That is, Dee's choice is dictated entirely by her constraints. So if choice is not anomalous, then no inferences can be made about Dee's preferences using rationalization, rational short list, limited attention, or limited consideration theory.

When choice is not anomalous, the idea of selecting a minimum constraint model can be combined with any aforementioned theory to identify preferences. Thus, the idea of minimum constraint can be used to select models within a given theory, but it does not screen among theories because, in principle at least, it can be combined with different theories. However, the basis of minimum constraint theory, permissibility assumption, that have been, so far, used to screen models within a given theory can also be used to screen among different theories.

A permissibility assumption has different interpretations depending on the associated theory. In limited consideration theory, $y \in B$ means that Dee is assumed to have considered option $y$ in issue $B$. In rationalization theory, $y \in B$ means that Dee is assumed to have a way to rationalize option $y$ in issue $B$. There may be evidence to substantiate the assumption that Dee has considered option $y$ (e.g., Dee talked about option $y$ right before making her choice), but not enough evidence to substantiate the assumption that Dee can rationalize option $y$ (e.g., nothing that Dee said about $y$ suggests it). If a permissibility assumption is made in one theory but not in another, then it is possible to empirically differentiate the two theories even if, in the absence of such assumptions, they have the same abstract structure as in the case of rationalization theory and limited consideration theory.

Consider the Thomas Schelling anecdote mentioned in Section 5.1. If Schelling is telling the story, it is sensible to assume that he was aware of the encyclopedias (and, naturally, the option of not buying them as well). This can be formalized by the permissibility assumptions such as $n \in (e_1, n)$ and $e_1 \in (e_1, e_2, n)$. Under these permissibility assumptions, the anomalous choices $C(e_1, n) = e_1$ (i.e., buying encyclopedia 1 when it is the only one offered) and $C(e_1, e_2, n) = n$ (i.e., not buying an encyclopedia when both encyclopedias are offered) lead to the contradictory conclusion that $e_1$ is preferred to $n$ and $n$ is preferred to $e_1$. So if these permissibility assumptions are made for limited consideration theory, but not for rationalization theory, then the two theories can be empirically differentiated.

The example above shows the use of nonchoice data, modeled by permissibility assumptions, to screen theories (see Kreps 1990 and Dekel and Lipman 2010 for a general discussion on nonchoice data). Additional examples readily apply. Nonchoice evidence may underlie a permissibility assumption for rationalization, but not for categorization. Assume that Dee said that "people should help the needy." This may substantiate the permissibility assumption that Dee can rationalize helping the needy (e.g., Dee can rationalize making a small donation), but may not substantiate any assumption regarding Dee's categorizations. Thus, although the categorize-then-choose theory has the same empirical scope as basic rationalization theory (and so choice data alone cannot

set them apart), these theories can be empirically distinguished when nonchoice data are added.

Consider the marketing field study of Berger and Smith (1997). They observe that some donors (to universities) choose to make a small solicited contribution ($s$) over no contribution ($n$), but if donors are solicited to make either a small or a large contribution ($l$), then they choose not to contribute. These two choices, $C(s,n)=s$ and $C(s,n,l)=n$, are anomalous. Depending on what is chosen between $n$ and $l$, and also between $s$ and $l$, we may end up with a cycle or an anomaly known in the literature as the *attraction effect*. Both patterns can be accommodated by order rationalization theory. However, regardless of what the two unobserved choices might be, by Proposition 2, Dee must prefer a small contribution over no contribution. Consider the permissibility assumption that Dee can rationalize a small donation (i.e., $s \in (s,n,l)$). This contradicts her choice of no donation. So under this assumption, rationalization should not be considered a viable explanation for this phenomenon. That is, rationalization theory is a poor candidate to model this phenomenon because the permissibility assumption that Dee can rationalize a small donation is plausible. So permissibility assumptions not only help select among alternative models of rationalization, but may also help circumscribe the application of the theory itself. We refer the reader to de Clippel and Eliaz (2012), Cherepanov et al. (forthcoming), Manzini and Mariotti (2012a), and Ok et al. (2012) for theories, among many others, that can accommodate the attraction effect. See also Masatlioglu and Nakajima (2007), Masatlioglu and Ok (2006), and Eliaz and Ok (2006) for related models.

Rubinstein and Salant (2006a) propose a postdominance 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 $R$. Rubinstein and Salant (2006b) show that postdominance rationality is characterized by an axiom called Exclusion Consistency. If Exclusion Consistency holds, then so does Weak WARP (a proof is available from the authors on request). Hence, the empirical scope of postdominance rationality theory is subsumed by the empirical scope of basic rationalization theory.

The literature also contains theories of multiple selves, where Dee always optimizes an order among several possible orders. These theories may accommodate anomalies and can also be empirically distinguished from rationalization theory. Consider the dual-self theory (a special case of Kalai et al. 2002), where Dee optimizes either by preference order $P_1$ or by preference order $P_2$. No other restriction is imposed. Consider four alternatives $x$, $y$, $z$, and $w$ and Dee's choices $C(x,y,z,w)=C(x,y)=x$ and $C(x,y,z)=y$. These choices violate Weak WARP and, therefore, cannot be accommodated by rationalization theory. However, they can be accommodated by dual-self theory. It suffice that one order ranks $y$ as a top option and another order ranks $x$ as a top option. Finally, we point out that the literature has already made an effort to relate different theories. For example, Houy (2007) and Houy and Tadenuma (2009) relate multiple-selves theories and theories of subjective constraints, Apesteguia and Ballester (2008) relate the Xu and Zhou (2007) theory of rationalizability by game tree with the work of Manzini and

Mariotti (2007), and Manzini and Mariotti (2012b) connect the Tversky (1969) model of boundedly rational choice to the models of choice of Apesteguia and Ballester (2008) and Manzini and Mariotti (2007). Finally, we point out that we have not provided a comprehensive survey of the relevant literature. In particular, we have not commented on related models such as Bossert and Suzumura (2009), Ergin and Saver (2010), and Salant and Rubinstein (2008), among many others.

### 6.1 *Future work*

Consider the following example (provided by an anonymous referee). Take the choices $\bar{C}$ in the handicapped aversion example, but add a fourth alternative $w$, where Dee sees movie 1 with her favorite person. She may choose $x$ (movie 1 alone) in the issue $(x, y, z, w)$ if now $x$ is psychologically permissible. Her choices violate Weak WARP and so cannot be accommodated by rationalization theory. Thus, it is desirable to produce a more flexible theory of choice, which is still testable and allows for some identification of its core elements.

In future research, rationalization may be used in game-theoretic models. For example, Dee's rationales may be, in part, social constructs. Then moral speech may affect rationales and, therefore, alter behavior.[7]

Future work may also shed light on welfare analysis when agents face psychological constraints. Suppose that Dee wants a life-saving medical procedure, but would choose against it because of a moral prohibition. Should someone acting on her behalf choose according to her preferences or according to her choice? Different perspectives have been offered on related matters (see, for example, Mill 1860 and Thaler and Sunstein 2003; see also Bernheim and Rangel 2009, Green and Hojman 2007, and Rubinstein and Salant 2012, among others, for a recent debate on welfare analysis). Rationalization theory can reveal Dee's preference and constraints and, hence, determine when they clash, but the welfare implications of such clashes are unresolved.

## 7. Conclusion

The inability to rationalize may place unobservable psychological constraints on choice. Rationalization theory imposes logical structure on psychological constraints and, thereby, guarantees that the theory is testable. Under minimum constraint rationalization theory, preferences and constraints are uniquely revealed across several choice patterns. When observed choice is not anomalous, minimum constraint theory reveals the same preferences as standard economics. When binary choice behavior is anomalous but acyclic, unique preferences and constraints are revealed from choice. When ambiguity over preferences remains, evidence that behavior is permissible may be used to reduce ambiguity and to reject the model outright. By combining the psychological idea of rationalization with the economic idea of ordered preferences and constrained choice, we get a new theory that can extend analysis in both disciplines.

[7]Simple game-theoretic examples where players are rationalizers are available on request. In some of these examples, rationalization theory produces behavior that resembles reciprocity (see Rabin 1993, Fehr and Schmidt 1999, and Bolton and Ockenfels 2000 for models of reciprocity).

While we consider a decision-theoretic framework, the rationalization model can serve as a foundation for strategic analyses. In particular, rationalization may help us to understand why debates about seemingly abstract principles might become a central feature of social life: such debates can change behavior without changing preferences or the feasibility of choice.

## Appendix

### A.1 *The need for orders*

In this section, we show that the handicapped avoidance claim cannot be established by basic rationalization theory and *any* permissibility assumptions. It is also necessary to assume that Dee's preferences are orders.

Given a choice function $C$, let $P^C$ be the binary relation defined by the binary choices. That is, $x\ P^C\ y$ if and only if $C(\{x, y\}) = x$.

Proposition 5. *Consider a model $(P, \psi) \in \mathcal{P} \times \bar{\Psi}$ that underlies a choice function $C$. Then the model $(P^C, \psi)$ also underlies the choice function $C$.*

Proposition 5 shows that if a choice function is consistent with basic rationalization theory, then it is always possible to accommodate Dee's choices by preferences defined by her binary choices. So in the handicapped avoidance example, there is a third way (i.e., beyond models $S$ and $S'$) to accommodate the choice function $\bar{C}$: with cyclic preferences $P^{\bar{C}}$ and the same psychological constraints in the $S$-model or in the $S'$-model. The intuition is as follows: Dee prefers her choice $C(B)$ over any rationalizable option $z \in \psi(B)$. In a binary choice between $z$ and $C(B)$, both options are rationalizable. Thus, Dee chooses $C(B)$ over $z$ in a binary choice.

In Proposition 5, the same psychological constraints $\psi$ are used in models $(P, \psi)$ and $(P^C, \psi)$ that accommodate choices $C$. This leads to Corollary 1 below. Let $\mathcal{P} \times \Psi^{\mathcal{A}}$ be the *basic theory of $\mathcal{A}$-rationalization.*

Corollary 1. *Consider the $\mathcal{P} \times \Psi^{\mathcal{A}}$ theory of $\mathcal{A}$-rationalization and a choice function $C$ such that $C(\{x, y\}) = x$. Then $y$ is not revealed to be preferred to $x$.*

Corollary 1 implies that it is not possible to infer that Dee acted against her preferences in a binary choice unless binary choice is cyclic and cyclic preferences are ruled out. This holds for any permissibility assumptions. Thus, the handicapped aversion claim *requires* cyclic choices *and* the assumption of preference orders.

### A.2 *An alternative theory*

Consider an alternative theory (suggested by an anonymous referee) in which given an issue $B \in \mathcal{B}$, Dee's constraint $\psi_{\mathcal{N}}(B)$ is defined by

$$\psi_{\mathcal{N}}(B) = \{x \in B \mid \text{if for all } y \in B,\ y \neq x, \text{ there is } R_i \in \mathcal{N} \text{ and } z \in B \text{ s.t. } z\ R_i\ y\},$$

where $\mathcal{N}$ is a finite collection of rationales. Informally, $x \in \psi_{\mathcal{N}}(B)$ if Dee can find a reason not to choose any feasible alternative.

This theory differs from order rationalization theory. Let $A$ be $x$, $y$, and $z$. Consider the constraint functions $\psi^1$ defined by $\psi^1(x,z) = (z)$ and $\psi^1(B) = B$ for every issue $B \neq (x,z)$, and $\psi^2$ defined by $\psi^2(B) = (B)$ for all binary issues and $\psi^2(x,y,z) = \{x\}$. It is straightforward to show that $\psi^1$ does not satisfy (1), but $\psi^1 = \psi_{\mathcal{N}}$ for some $\mathcal{N}$, and $\psi^2$ satisfies (1), but there is no $\mathcal{N}$ such that $\psi^2 = \psi_{\mathcal{N}}$.

### A.3 *Proofs and extended results*

Given a set of rationales $\mathcal{R} = \{R_i, i = 1, \ldots, n\}$, let $\tilde{\psi}^{\mathcal{R}}$ be the psychological constraint function

$$\tilde{\psi}^{\mathcal{R}}(B) = \{x \in B \mid \text{for some } R_i \in \mathcal{R} \text{ there is no option } y \neq x,\ y \in B \text{ s.t. } y\, R_i\, x\}.$$

PROPOSITION 7. *Let $\psi \in \Psi$ be any psychological constraint function. The following conditions are equivalent.*

*(i) The mapping $\psi$ satisfies (1).*

*(ii) There exists a set $\mathcal{R}$ of rationales (where each rationale in $\mathcal{R}$ is transitive and asymmetric) such that $\psi = \psi^{\mathcal{R}}$.*

*(iii) There exists a set $\mathcal{R}$ of rationales (where each rationale in $\mathcal{R}$ is transitive and asymmetric) such that $\psi = \tilde{\psi}^{\mathcal{R}}$.*

PROOF. (i) $\Longrightarrow$ (ii) Assume that a psychological constraint function $\psi$ satisfies (1). Then, for each issue $B \in \mathcal{B}$ and alternative $x \in \psi(B)$, let $R_{B,x}$ be defined by $x\, R_{B,x}\, y$ for any $y \in B$, $y \neq x$. So $x\, R_{B,x}\, y$ if and only if $x \in \psi(B)$, $y \in B$, and $y \neq x$. Let $\mathcal{R}$ be the set of all rationales $R_{B,x}$ such that $B \in \mathcal{B}$ and $x \in \psi(B)$. Let $\psi^{\mathcal{R}}$ be the psychological constraint function determined by $\mathcal{R}$. Fix any issue $B \in \mathcal{B}$. Assume that $x \in \psi(B)$. Then, by definition, $x$ is rationalized by $R_{B,x} \in \mathcal{R}$. So $x \in \psi^{\mathcal{R}}(B)$. Now assume that $x \in \psi^{\mathcal{R}}(B)$. So $x \in B$ and there exists an issue $\tilde{B}$ such that $x\, R_{\tilde{B},x}\, y$ for any $y \in B$, $y \neq x$. By definition, $x\, R_{\tilde{B},x}\, y$ if and only if $x \in \psi(\tilde{B})$, $y \in \tilde{B}$, and $y \neq x$. So $x \in \psi(\tilde{B})$. By (1), $x \in \psi(B)$.

(ii) $\Longrightarrow$ (iii) is immediate. (iii) $\Longrightarrow$ (i) can be shown as follows. Assume that $x \in B \subseteq B^*$ and $x \in \tilde{\psi}^{\mathcal{R}}(B^*)$. Then, by definition, there is some $R_i \in \mathcal{R}$ such that there is no alternative $y \neq x$, $y \in B^*$ such that $y\, R_i\, x$. Hence, there is no alternative $y \neq x$, $y \in B$ such that $y\, R_i\, x$. So $x \in \tilde{\psi}^{\mathcal{R}}(B)$. □

The equivalence between (i) and (ii) shows that (1) is the structure on psychological constraint functions imposed by rationalization. It also shows that, without loss of generality, rationales can be transitive and asymmetric. The equivalence between (ii) and (iii) shows that our results are the same whether an option is rationalized when it is ranked highest for some rationale or whether some rationale does not place any alternative above $x$. Finally, it is not without loss of generality to assume that rationales are

orders. Consider the psychological constraint $\psi'\{y,z\} = \psi'\{x,y,z\} = \{y\}$, $\psi'\{B\} = B$ elsewhere. If $x \in \psi'\{x,y\}$ and rationales are orders, then Dee must have an ordered rationale that ranks $x$ above $y$. The highest ranked option in this rationale is not $y$. This contradicts $\psi'\{x,y,z\} = \{y\}$. Note that $\psi' \in \bar{\Psi}$ is the psychological constraint function in the $S'$-model that accommodates $\bar{C}$ even though it does not involve handicapped aversion. If rationales are orders, then $\bar{C}$ cannot be accommodated without handicapped aversion. By Proposition 2, $x$ is revealed to be preferred to $y$, so if $z$ is preferred to $x$ (i.e., no handicap aversion), then Dee ranks $z$ above $x$ above $y$. To accommodate $\bar{C}(x,y,z) = y$ requires $\psi\{x,y,z\} = \{y\}$ and to accommodate $\bar{C}(x,y) = x$ requires $x \in \psi\{x,y\}$. This leads to the contradiction obtained above. So if rationales are orders, then handicapped aversion follows from $\bar{C}$ without permissibility assumptions.

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*. Given a choice function $C$, a pair of nested issues $(B, B^*) \in \mathcal{B} \times \mathcal{B}$ is anomalous if the choices $C(B)$ and $C(B^*)$ are anomalous. Given an issue $B$, let $\mathcal{B}^B$ be all super-issues $B^*$ of $B$ such that the pair $(B, B^*)$ is anomalous. Given a choice function $C$ and a set $\mathcal{A} = \{(y_i, B_i);\, y_i \in B_i \; i = 1, \ldots, n\}$, let $P^{C,\mathcal{A}}$ be the binary relation such that $x\, P^{C,\mathcal{A}}\, y$ if and only if

$$x = C(B) \quad \text{and} \quad y = C\big(B^*\big) \quad \text{for some anomalous pair } \big(B, B^*\big) \text{ of nested issues}$$

or

$$x = C(B) \quad \text{and} \quad \text{for some } (y_i, B_i) \in \mathcal{A}, \quad y = y_i \in B \subseteq B_i.$$

Let $\psi^{C,\mathcal{A}}$ be a psychological constraint function defined by

$$\psi(B) = \big\{C(B);\, C\big(B^*\big) \text{ for any } B^* \in \mathcal{B}^B;\, y_i \text{ for any } (y_i, B_i) \in \mathcal{A},\, y = y_i \in B \subseteq B_i\big\}.$$

By definition,

$$C(B)\, P^{C,\mathcal{A}}\, y \quad \text{for any } y \in \psi^{C,\mathcal{A}}(B),\, y \neq C(B). \tag{4}$$

In addition, if $B \subseteq \tilde{B}$, then

$$\psi^{C,\mathcal{A}}(\tilde{B}) \cap B \subseteq \psi^{C,\mathcal{A}}(B).$$

This follows because if $z \in B$ and $z \in \psi^{C,\mathcal{A}}(\tilde{B})$, then we can assume, without loss of generality, that $z \neq C(B)$. Otherwise $z = C(B)$ and so $z \in \psi^{C,\mathcal{A}}(B)$. We can also assume, without loss of generality, that $z \neq C(\tilde{B})$ and that $z \neq C(\hat{B})$ for any $\hat{B} \in \mathcal{B}^{\tilde{B}}$. Otherwise $(B, \tilde{B})$ or $(B, \hat{B})$ is an anomalous pair of nested issues and in either case, $z \in \psi^{C,\mathcal{A}}(B)$. Thus, it follows from $z \in \psi^{C,\mathcal{A}}(\tilde{B})$ that for some $(y_i, B_i) \in \mathcal{A}$, $z = y_i \in \tilde{B} \subseteq B_i$. So $z = y_i \in B \subseteq \tilde{B} \subseteq B_i$. Thus, $z \in \psi^{C,\mathcal{A}}(B)$. So $\psi^{C,\mathcal{A}} \in \Psi^{\mathcal{A}}$.

Lemma 1. *If $(P, \psi) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ underlies $C$, then $x\, P^{C,\mathcal{A}}\, y \Longrightarrow x\, P\, y$.*

Proof. Assume that $x = C(B)$ and $y = C(B^*)$ for some anomalous pair $(B, B^*)$ of nested issues. Then $y \in \psi(B^*)$ (because $y = C(B^*)$) and $y \in B$ (because $C(B^*) \in B$). So by (1),

$y \in \psi(B)$. Hence, $x\ P\ y$ (because $(P, \psi)$ underlies $C$). Now assume that $x = C(B)$ and for some $(y_i, B_i) \in \mathcal{A}$, $y = y_i \in B \subseteq B_i$. So $y_i \in \psi(B_i)$ (because $\psi \in \Psi^{\mathcal{A}}$). By (1), $y_i \in \psi(B)$. Hence, $x\ P\ y = y_i$. □

PROPOSITION A.1. *A choice function $C$ is consistent with $\mathcal{A}$-rationalization theory $\mathcal{P} \times \Psi^{\mathcal{A}}$ if and only if $P^{C,\mathcal{A}}$ is asymmetric. A choice function $C$ is consistent with $\mathcal{A}$-rationalization order theory $\mathcal{P}^o \times \Psi^{\mathcal{A}}$ if and only if $P^{C,\mathcal{A}}$ is acyclic.*

PROOF. Assume that a choice function $C$ is consistent with $\mathcal{A}$-rationalization theory $\mathcal{P} \times \Psi^{\mathcal{A}}$. Let $(P, \psi) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ be a model that underlies $C$. Assume, to the contrary, that $P^{C,\mathcal{A}}$ is not asymmetric. Then, for some $x \neq y$, $x\ P^{C,\mathcal{A}}\ y$ and $y\ P^{C,\mathcal{A}}\ x$. By Lemma 1, $x\ P\ y$ and $y\ P\ x$. This contradicts $P \in \mathcal{P}$. Now assume that $P^{C,\mathcal{A}}$ is asymmetric. Then, by (4), $(P^{C,\mathcal{A}}, \psi^{C,\mathcal{A}}) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ underlies $C$.

Assume that a choice function $C$ is consistent with order $\mathcal{A}$-rationalization theory $\mathcal{P}^o \times \Psi^{\mathcal{A}}$. Let $(P, \psi) \in \mathcal{P}^o \times \Psi^{\mathcal{A}}$ be a model that underlies $C$. Assume, to the contrary, that $P^{C,\mathcal{A}}$ is cyclic. By Lemma 1, $P$ is cyclic. This contradicts $P \in \mathcal{P}^o$. Assume that $P^{C,\mathcal{A}}$ is acyclic. By topological ordering, $P^{C,\mathcal{A}}$ may be extended (not necessarily uniquely) to an order (see Cormen et al. 2001, pp. 549–552). Let $\bar{P}$ be an arbitrary order that extends $P^{C,\mathcal{A}}$. Then, by (4), $(\bar{P}, \psi^{C,\mathcal{A}}) \in \mathcal{P}^o \times \Psi^{\mathcal{A}}$ underlies $C$. □

Proposition A.1 shows the empirical content of (order) $\mathcal{A}$-rationalization theory.

PROPOSITION A.2. *A choice function $C$ that is consistent with order rationalization theory is also consistent with the minimum constraint theory of rationalization.*

PROOF. Let us define the partial order $\geqq$ on a psychological constraint function such that $\psi' \geqq \psi$ if and only if $\psi(B) \subseteq \psi'(B)$ for all issues $B \in \mathcal{B}$. Given that the set of all alternatives $A$ is finite, there is a $\geqq$-maximal element, $\psi^*$, in the set of $\{\psi : \text{for some } P, (P, \psi) \in \mathcal{P}^o \times \Psi^{\mathcal{A}} \text{ underlies } C\}$. So, for some $P^*$, $(P^*, \psi^*) \in \mathcal{P}^o \times \Psi^{\mathcal{A}}$ underlies $C$ and is not dominated by any alternative model in $\mathcal{P}^o \times \Psi^{\mathcal{A}}$ that underlies $C$. □

PROPOSITION A.3. *Consider a choice function $C$ consistent with $\mathcal{A}$-rationalization theory $\mathcal{P} \times \Psi^{\mathcal{A}}$. Then $x$ is revealed to be preferred to $y$, $x \neq y$, if and only if at least one of the two conditions hold: (i) $x$ is revealed to be preferred to $y$ by basic rationalization theory or (ii) $x = C(B)$ and for some $(y_i, B_i) \in \mathcal{A}$, $y = y_i \in B \subseteq B_i$.*

PROOF. Let $(P, \psi) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ be a model that underlies $C$. So if either condition (i) or (ii) holds, then $x\ P^{C,\mathcal{A}}\ y$ and, by Lemma 1, $x\ P\ y$. Now assume that $C$ is consistent with $\mathcal{A}$-rationalization theory $\mathcal{P} \times \Psi^{\mathcal{A}}$. Then, by Proposition A.1, $P^{C,\mathcal{A}}$ is asymmetric. Hence, $(P^{C,\mathcal{A}}, \psi^{C,\varnothing}) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ underlies $C$. If neither condition (i) nor condition (ii) holds, then it is *not* the case that $x\ P^{C,\mathcal{A}}\ y$. Thus, consider the binary relation $\bar{P}$ such that $y\ \bar{P}\ x$ and for all other pairs of alternatives, $\bar{P}$ is identical to $P^{C,\mathcal{A}}$. Then $\bar{P}$ is still asymmetric and $(\bar{P}, \psi^{C,\mathcal{A}}) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ still underlies $C$. □

PROPOSITION A.4. *Consider a choice function $C$ consistent with order $\mathcal{A}$-rationalization theory $\mathcal{P}^o \times \Psi^{\mathcal{A}}$. Then $z_1$ is revealed to be preferred to $z_k$ if and only if there exists a chain $z_{i+1}$, $i = 0, \ldots, k-1$, such that $z_i$ is revealed to be preferred to $z_{i+1}$ by $\mathcal{A}$-rationalization theory.*

PROOF. Let $(P, \psi) \in \mathcal{P}^o \times \Psi^{\mathcal{A}}$ be a model that underlies $C$. If there exists a chain $z_{i+1}$, $i = 0, \ldots, k-1$, such that $z_i$ is revealed to be preferred to $z_{i+1}$ by $\mathcal{A}$-rationalization theory, then $z_i \, P \, z_{i+1}$, $i = 0, \ldots, k-1$, which implies (because $P$ is an order) that $z_1 \, P \, z_k$. Now assume that $C$ is consistent with $\mathcal{A}$-rationalization order theory $\mathcal{P}^o \times \Psi^{\mathcal{A}}$. By Proposition A.1, $P^{C,\mathcal{A}}$ is acyclic. Now assume that there is no chain $z_{i+1}$, $i = 0, \ldots, k-1$, such that $z_i$ is revealed to be preferred to $z_{i+1}$ by $\mathcal{A}$-rationalization theory. Then it is *not* the case that $z_1 \, P^{C,\mathcal{A}} \, z_k$. Thus, consider the binary relation $\bar{P}$ such that $z_k \, \bar{P} \, z_1$ and for all other pairs of alternatives, $\bar{P}$ is identical to $P^{C,\mathcal{A}}$. Then $\bar{P}$ is acyclic and, hence, can be extended to an order $\hat{P} \in \mathcal{P}^o$. Given that $\hat{R}$ extends $P^{C,\mathcal{A}}$ and that $(P^{C,\mathcal{A}}, \psi^{C,\mathcal{A}}) \in \mathcal{P} \times \Psi^{\mathcal{A}}$ underlies $C$, $(\hat{P}, \psi^{C,\mathcal{A}}) \in \mathcal{P}^o \times \Psi^{\mathcal{A}}$ underlies $C$. □

Propositions A.3 and A.4 characterizes the preference inferences that can made under (order) $\mathcal{A}$-rationalization theory. We now return to the basic rationalization theory (i.e., $\mathcal{A} = \varnothing$ and preferences are not necessarily orders). Consider two pairs of anomalous nested issues $(B_1, B_1^*)$ and $(B_2, B_2^*)$. The choices on these two nested issues are *reversed* if $C(B_1) = C(B_2^*)$ and $C(B_1^*) = C(B_2)$.

IRREVERSIBILITY. A choice function $C$ satisfies the Irreversibility axiom if there are no two pairs of anomalous nested issues with reversed choices.

By Proposition A.1, this axiom demarcates the choice functions that can be accommodated by the basic rationalization theory because there are two pairs of anomalous nested issues with reversed choices if and only if $P^{C,\varnothing}$ is asymmetric.

PROPOSITION A.5. *The Irreversibility axiom holds if and only if Weak WARP holds.*

PROOF. Assume that Weak WARP 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 anomalous nested issues and $(B, \bar{B})$ is also a pair of anomalous nested issues. 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 anomalous nested issues with reversed choices. 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 Weak WARP does not hold. Alternatively, if $C(\{x, y\}) = y$, then $\{x, y\} \subseteq B_2 \subseteq B_2^*$, $C(B_2^*) = y$, and $C(B_2) = x$. Thus, Weak WARP does not hold. □

The proof of Proposition 1 is a direct corollary of Propositions A.1 and A.5.

Proof of Proposition 2. Let $(P,\psi)\in\mathcal{P}\times\bar{\Psi}$ be a model that underlies $C$. Note that $\bar{\Psi}=\Psi^{\mathcal{A}}$, where $\mathcal{A}=\varnothing$. So if $(B,B^*)$ is an anomalous pair of nested issues, then $C(B)\,P^{C,\varnothing}\,C(B^*)$ and, by Lemma 1, $C(B)\,P\,C(B^*)$. So $C(B)$ is revealed to be preferred to $C(B^*)$. In addition, if $P$ is transitive, then $x_i\,P\,x_{i+1}$, $i=0,\dots,n-1$, implies that $x=x_0\,P\,x_n=y$.

If $C$ is consistent with rationalization theory $\mathcal{P}\times\bar{\Psi}$, then, by Proposition A.1, $P^{C,\varnothing}$ is asymmetric. Hence, $(P^{C,\varnothing},\psi^{C,\varnothing})\in\mathcal{P}\times\bar{\Psi}$ underlies $C$. If there exists no anomalous pair of nested issues $(B,B^*)$ such that $C(B)=x$ and $C(B^*)=y$, $x\neq y$, then it is *not* the case that $x\,P^{C,\varnothing}\,y$. Thus, consider the binary relation $P^a$ such that $y\,P^a\,x$ and for all other pairs of alternatives, $P^a$ is identical to $P^{C,\varnothing}$. Then $P^a$ is still asymmetric and $(P^a,\psi^{C,\varnothing})\in\mathcal{P}\times\bar{\Psi}$ still underlies $C$. In addition, if $C$ is consistent with rationalization theory $\mathcal{P}\times\bar{\Psi}$, then, by Proposition A.1, $P^{C,\varnothing}$ is acyclic. Therefore, if there exists no chain of alternatives $\{x_0,\dots,x_n\}$ such that $x_0=x$, $y=x_n$, and $x_i\,P^{C,\varnothing}\,x_{i+1}$, $i=0,\dots,n-1$, then $P^a$ (as defined above) remains acyclic. Hence, $P^a$ can be extended to an order $\bar{P}$ and, by construction, $(\bar{P},\psi^{C,\varnothing})\in\mathcal{P}\times\bar{\Psi}$ still underlies $C$. □

Proof of Theorem 1. Assume that an order $P$ is identifiable. Let $\bar{\psi}$ be a $\geqq$-maximal element in $\{\psi:(P,\psi)\in\mathcal{P}^o\times\Psi^{\mathcal{A}}\text{ underlies }C\}$. So $(P,\bar{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$ underlies $C$. Assume, to the contrary, that for $(P,\bar{\psi})$ there exists a model $(P',\psi')\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$ that also underlies $C$ such that $\bar{\psi}(B)\subseteq\psi'(B)$ for all issues $B\in\mathcal{B}$, with strict inclusion for some issue $B\in\mathcal{B}$. Then $P'\neq P$ ($P'=P$ contradicts the $\geqq$-maximality of $\bar{\psi}$). Consider the model $(P',\bar{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$. First, $(P',\bar{\psi})$ also underlies $C$, because for any issue $B$, $C(B)\,P'\,y$ for all $y\in\psi'(B)\supseteq\bar{\psi}(B)$. So $C(B)\,P'\,y$ for all $y\in\bar{\psi}(B)$. By definition, for any permissibility assumptions $\mathcal{A}$, if $(P,\bar{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$, then $(P',\bar{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$. Thus, $P$ is not identified by $\mathcal{P}^o\times\Psi^{\mathcal{A}}$.

Assume that $P$ is an order and that $(P,\psi)$ belongs to the minimum constraint theory of rationalization. Let $\mathcal{A}$ be permissibility assumptions defined $(y,B)\in\mathcal{A}$ if and only if $y\in\psi(B)$. Assume, to the contrary, that $P$ is not identified by $\mathcal{P}^o\times\Psi^{\mathcal{A}}$. Then there exists a model $(P',\psi')\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$ that underlies $C$, with $P'\neq P$. Now $\psi'\geqq\psi$ (because $\psi'\in\Psi^{\mathcal{A}}$). So $\psi'=\psi$ (otherwise $(P,\psi)$ is dominated by $(P',\psi')$). Thus, $(P',\psi)\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$ underlies $C$. Now let $a$ and $b$ be two alternatives such that $a\,P\,b$ and $b\,P'\,a$ (they exist because $P'\neq P$). Let $\hat{\psi}$ be identical to $\psi$ on all issues $B\neq\{a,b\}$ and $\hat{\psi}\{a,b\}=\{a,b\}$. By definition, if $\{a,b\}\subseteq\tilde{B}$, then $\hat{\psi}(\tilde{B})\cap\{a,b\}\subseteq\hat{\psi}\{a,b\}$ and $y\in\hat{\psi}(B)$ if $(y,B)\in\mathcal{A}$. Thus, $\hat{\psi}\in\Psi^{\mathcal{A}}$. Now either $b\in\psi\{a,b\}$ or $b\notin\psi\{a,b\}$. In the latter case, $\psi\{a,b\}=\{a\}$, $C(a,b)=a$, and $(P,\hat{\psi})$ underlies $C$ (because $a\,P\,b$). Thus, $(P,\psi)$ is dominated by $(P,\hat{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$—a contradiction. In the former case, $b\in\psi\{a,b\}$. Then $C\{a,b\}=b$ (because $b\,P'\,a$ and $(P',\psi)$ underlies $C$). Thus, $(P',\hat{\psi})$ underlies $C$. In addition, $\psi\{a,b\}=\{b\}$ ($\psi\{a,b\}=\{a,b\}$ would contradict $a\,P\,b$ and $(P,\psi)$ underlies $C$). Hence, $(P,\psi)$ is dominated by $(P',\hat{\psi})\in\mathcal{P}^o\times\Psi^{\mathcal{A}}$—a contradiction. □

Proof of Proposition 3. Assume that $(P,\psi)\in\mathcal{P}\times\bar{\Psi}$ underlies $C$ and is not dominated by any alternative model in $\mathcal{P}^o\times\bar{\Psi}$. Then, for every pair $\{x,y\}\subset A$, $\psi\{x,y\}=\{x,y\}$. To see this, assume, to the contrary, that for some pair of alternatives $\{x,y\}$, $\psi\{x,y\}=\{x\}$. Let $\psi'$ be such that $\psi'\{x,y\}=\{x,y\}$ and $\psi'=\psi$ for all other issues. Clearly, $\psi'\in\bar{\Psi}$ because

$\psi \in \bar{\Psi}$ and $\{x, y\}$ has no sub-issues with more than one alternative. By assumption, $P^C$ (defined in Appendix A.1) is complete and acyclical, and so is an order. We now show that $(P^C, \psi') \in \mathcal{P}^o \times \bar{\Psi}$ underlies $C$.

Let $B \neq \{x, y\}$ be an issue. Let $z \in \psi'(B) = \psi(B)$, $z \neq C(B)$. Note that $\{C(B), z\} \subseteq B$ and $C(B) \in \psi(B) = \psi'(B)$. So $\{C(B), z\} \subseteq B \cap \psi'(B)$ and $\{C(B), z\} \subseteq B \cap \psi(B)$. Hence, $\psi(\{C(B), z\}) = \psi'(\{C(B), z\}) = \{C(B), z\}$. It follows that $C(B)\ P\ z$ (because $z \in \psi(\{C(B), z\})$ and $(R, \psi)$ underlies $C$). Hence, $C(\{C(B), z\}) = C(B)$ (because $(R, \psi)$ underlies $C$). By definition, $C(B)\ P^C\ z$. Moreover, $C(\{x, y\}) = x$ (because $\psi\{x, y\} = \{x\}$). So $x\ P^C\ y$. Hence, $(P^C, \psi') \in \mathcal{P}^o \times \bar{\Psi}$ underlies $C$ and $(P, \psi)$ is dominated by $(P^C, \psi')$. Thus, for every pair of alternatives $\{x, y\}$, $\psi\{x, y\} = \{x, y\}$. Given that $(P, \psi)$ underlies $C$, it follows that $P = P^C$. □

Proof of Proposition 4. Let $(P, \psi)$ and $(P, \psi')$ be two models in $\mathcal{P}^o \times \bar{\Psi}$ that underlie choices $C$ and are not dominated by any alternative model in $\mathcal{P}^o \times \bar{\Psi}$. Assume, to the contrary, that some issue $\bar{B}$, $\psi(\bar{B}) \neq \psi'(\bar{B})$. Let $\hat{\psi}$ be defined by $\hat{\psi}(B) = \psi(B) \cup \psi'(B)$. By definition, either $\psi$ or $\psi'$ (or both) imposes more constraints than $\hat{\psi}$. Now $(P, \hat{\psi})$ underlies $C$ because $C(B)\ P\ y$ for every $y \in \psi(B)$ and for every $y \in \psi'(B)$. In addition, if $B \subseteq \tilde{B}$, then $\hat{\psi}(\tilde{B}) \cap B \subseteq \hat{\psi}(B)$. This follows because $\psi(\tilde{B}) \cap B \subseteq \psi(B)$ and $\psi'(\tilde{B}) \cap B \subseteq \psi'(B)$. So $\hat{\psi} \in \bar{\Psi}$. Thus, either $(P, \psi)$ or $(P, \psi')$, or both, are dominated by $(P, \hat{\psi})$—a contradiction. □

Proof of Proposition 5. Let $(P, \psi)$ underlie $C$. Fix an issue $B \in \mathcal{B}$ and an alternative $z \in \psi(B)$. Now $C(B) \in \psi(B)$ and $z \in \psi(B)$ implies that $\{C(B), z\} \subseteq B \cap \psi(B)$. Therefore, $\psi\{C(B), z\} = \{C(B), z\}$. Since $C(B)\ P\ z$ (because $(P, \psi)$ underlies $C$), it must be the case that $C(\{C(B), z\}) = C(B)$. Thus, $C(B)\ P^C\ z$. □

Proof of the claim in Section 6. Let $\psi^{\mathrm{SL}}$ be a constraint function in Manzini and Mariotti (2007). Let $P_1$ be the associated preference such that (2) is satisfied. Let $\mathcal{R} = \{P^x, x \in A\}$ be the set of all rationales defined by $x\ P^x\ y$ whenever $y\ P_1\ x$ does not hold, $y \neq x$. Then $\psi^{\mathrm{SL}}(B) = \psi^{\mathcal{R}}(B)$ for all $B \in \mathcal{B}$. So $\psi^{\mathrm{SL}} \in \bar{\Psi}$.

Now assume that $A$ has three elements $x$, $y$, and $z$. Let $\psi(B) = B$ for all binary issues and let $\psi(x, y, z) = (x, z)$. Assume, to the contrary, that $\psi$ satisfies (2). Then the associated preference $P_1$ cannot rank any alternative above another (because $\psi(B) = B$ for all binary issues). But this implies $\psi(x, y, z) = (x, y, z)$—a contradiction. Thus, (2) is not satisfied. It is immediate that $\psi$ satisfies (1). □

## References

Achen, Christopher H. and Larry M. Bartels (2006), "It feels like we're thinking: The rationalizing voter and electoral democracy." Unpublished paper. [778]

Akerlof, George A. and William T. Dickens (1982), "The economic consequences of cognitive dissonance." *American Economic Review*, 72, 307–319. [778]

Ambrus, Attila and Kareen Rozen (2008), "Revealed conflicting preferences." Unpublished paper. [777]

Apesteguia, Jose and Miguel Ballester (2008), "A characterization of sequential rationalizability." Unpublished paper. [788, 789]

Bénabou, Roland and Jean Tirole (2002), "Self-confidence and personal motivation." *Quarterly Journal of Economics*, 117, 871–915. [778]

Berger, Paul D. and Gerald E. Smith (1997), "The effect of direct mail framing strategies and segmentation variables on university fundraising performance." *Journal of Interactive Marketing*, 11, 30–43. [788]

Bernheim, B. Douglas (1984), "Rationalizable strategic behavior." *Econometrica*, 52, 1007–1028. [778]

Bernheim, B. Douglas and Antonio Rangel (2009), "Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics." *Quarterly Journal of Economics*, 124, 51–104. [789]

Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 90, 166–193. [789]

Bossert, Walter and Kotaro Suzumura (2009), "External norms and rationality of choice." *Economics and Philosophy*, 25, 139–152. [789]

Carillo, Juan D. and Thomas Mariotti (2000), "Strategic ignorance as a self-disciplining device." *Review of Economic Studies*, 67, 529–544. [778]

Chambers, Christopher P. and Takashi Hayashi (2012), "Choice and individual welfare." *Journal of Economic Theory*, 147, 1818–1849. [777]

Chen, M. Keith (2008), "Rationalization and cognitive dissonance: Do choices affect or reflect preferences?" Unpublished paper. [778]

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni (forthcoming), "Revealed preferences and aspirations in warm glow theory." *Economic Theory*. [788]

Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein (2001), *Introduction to Algorithms*, second edition. McGraw-Hill, New York. [793]

de Clippel, Geoffroy and Kfir Eliaz (2012), "Reason-based choice: A bargaining rationale for the attraction and compromise effects." *Theoretical Economics*, 7, 125–162. [777, 788]

Dekel, Eddie and Barton L. Lipman (2010), "How (not) to do decision theory." *Annual Review of Economics*, 2, 257–282. [787]

Dietrich, Franz and Christian List (2010), "A reason-based theory of rational choice." Unpublished paper. [778]

Ehlers, Lars and Yves Sprumont (2008), "Weakened WARP and top-cycle choice rules." *Journal of Mathematical Economics*, 44, 87–94. [779]

Eliaz, Kfir, Michael Richter, and Ariel Rubinstein (2011), "Choosing the two finalists." *Economic Theory*, 46, 211–219. [785]

Eliaz, Kfir and Efe Ok (2006), "Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences." *Games and Economic Behavior*, 56, 61–86. [778, 788]

Eliaz, Kfir and Ran Spiegler (2011), "Consideration sets and competitive marketing." *Review of Economic Studies*, 78, 235–262. [785]

Ergin, Haluk and Todd Saver (2010), "A unique costly contemplation representation." *Econometrica*, 78, 1285–1339. [789]

Fehr, Ernst and Klaus M. Schmidt (1999), "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114, 817–868. [789]

Fudenberg, Drew and David K. Levine (2006), "A dual-self model of impulse control." *American Economic Review*, 96, 1449–1476. [778]

Gneezy, Uri and Aldo Rustichini (2000), "A fine is a price." *Journal of Legal Studies*, 29, 1–17. [778]

Green, Jerry R. and Daniel A. Hojman (2007), "Choice, rationality and welfare measurement." Unpublished paper. [789]

Gul, Faruk and Wolfgang Pesendorfer (2005), "The revealed preference theory of changing tastes." *Review of Economic Studies*, 72, 429–448. [778]

Heller, Yuval (2012), "Justifiable choice." *Games and Economic Behavior*, 76, 375–390. [778]

Houy, Nicolas (2007), "Rationality and order-dependent sequential rationality." *Theory and Decision*, 62, 119–134. [788]

Houy, Nicolas and Koichi Tadenuma (2009), "Lexicographic compositions of multiple criteria for decision making." *Journal of Economic Theory*, 144, 1770–1782. [788]

Iyengar, Sheena S. and Mark R. Lepper (2000), "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology*, 79, 995–1006. [784]

Jones, Ernest (1908), "Rationalisation in every-day life." *Journal of Abnormal Psychology*, 3, 161–169. [775]

Kalai, Gil, Ariel Rubinstein, and Ran Spiegler (2002), "Rationalizing choice functions by multiple rationales." *Econometrica*, 70, 2481–2488. [778, 788]

Kreps, David M. (1990), *A Course in Microeconomic Theory*. Princeton University Press, Princeton, New Jersey. [787]

Lehrer, Ehud and Roee Teper (2011), "Justifiable preferences." *Journal of Economic Theory*, 146, 762–774. [778]

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Ozbay (2010), "When more is less: Limited consideration." Unpublished paper. [778, 785, 786]

Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Ozbay (2012), "When more is less: Limited consideration." Unpublished paper. [786]

Manzini, Paola and Marco Mariotti (2007), "Sequentially rationalizable choice." *American Economic Review*, 97, 1824–1839. [778, 779, 785, 788, 789, 796]

Manzini, Paola and Marco Mariotti (2012a), "Categorize then choose: Boundedly rational choice and welfare." *Journal of the European Economic Association*, 10, 1141–1165. [778, 779, 785, 786, 788]

Manzini, Paola and Marco Mariotti (2012b), "Choice by lexicographic semiorders." *Theoretical Economics*, 7, 1–23. [789]

Masatlioglu, Yusufcan and Daisuke Nakajima (2007), "A theory of choice by elimination." Unpublished paper. [778, 785, 788]

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012), "Revealed attention." *American Economic Review*, 102, 2183–2205. [778, 785, 786]

Masatlioglu, Yusufcan and Efe Ok (2006), "Reference-dependent procedural decision making." Unpublished paper. [788]

Mazar, Nina and Dan Ariely (2006), "Dishonesty in everyday life and its policy implications." *Journal of Public Policy and Marketing*, 25, 117–126. [778]

Mill, John Stuart (1860), *On Liberty*. Collier, London. [789]

Ok, Efe A., Pietro Ortoleva, and Gil Riella (2012), "Revealed (p)reference theory." Unpublished paper. [778, 788]

Pearce, David G. (1984), "Rationalizable strategic behavior and the problem of perfection." *Econometrica*, 52, 1029–1050. [778]

Rabin, Matthew (1993), "Incorporating fairness into game theory and economics." *American Economic Review*, 83, 1281–1302. [789]

Rabin, Matthew (1995), "Moral preferences, moral constraints, and self-serving biases." Unpublished paper. [778]

Roth, Alvin E. (2007), "Repugnance as a constraint on markets." Unpublished paper. [778]

Rubinstein, Ariel and Yuval Salant (2006a), "A model of choice from lists." *Theoretical Economics*, 1, 3–17. [778, 788]

Rubinstein, Ariel and Yuval Salant (2006b), "Two comments on the principle of revealed preference." Unpublished paper. [788]

Rubinstein, Ariel and Yuval Salant (2012), "Eliciting welfare preferences from behavioural data sets." *Review of Economic Studies*, 79, 375–387. [789]

Salant, Yuval and Ariel Rubinstein (2008), "$(a, f)$: Choice with frames." *Review of Economic Studies*, 75, 1287–1296. [789]

Samuelson, Paul A. (1938), "The empirical implications of utility analysis." *Econometrica*, 6, 344–356. [776]

Sen, Amartya (1997), "Maximization and the act of choice." *Econometrica*, 65, 745–779. [781]

Shafir, Eldar, Itamar Simonson, and Amos Tversky (1993), "Reason-based choice." *Cognition*, 49, 11–36. [784]

Snyder, Melvin L., Robert E. Kleck, Angelo Strenta, and Steven J. Mentzer (1979), "Avoidance of the handicapped: An attributional ambiguity analysis." *Journal of Personality and Social Psychology*, 37, 2297–2306. [776, 781]

Spiegler, Ran (2002), "Equilibrium in justifiable strategies: A model of reason-based choice in extensive-form games." *Review of Economic Studies*, 69, 691–706. [778]

Spiegler, Ran (2004), "Simplicity of beliefs and delay tactics in a concession game." *Games and Economic Behavior*, 47, 200–220. [778]

Sprumont, Yves (2000), "On the testable implications of collective choice theories." *Journal of Economic Theory*, 93, 205–232. [778]

Thaler, Richard H. and Cass R. Sunstein (2003), "Libertarian paternalism." *American Economic Review: Papers and Proceedings*, 93, 175–179. [789]

Tversky, Amos (1969), "Intransitivity of preferences." *Psychological Review*, 76, 31–48. [789]

Tversky, Amos and Eldar Shafir (1992), "Choice under conflict: The dynamics of deferred decision." *Psychological Science*, 3, 358–361. [784]

Tversky, Amos and Itamar Simonson (1993), "Context-dependent preferences." *Management Science*, 39, 1179–1189. [784]

von Hippel, William, Jessica Lakin, and Richard Shakarchi (2005), "Individual differences in motivated social cognition: The case of self-serving information processing." *Personality and Social Psychology Bulletin*, 31, 1347–1357. [778]

Xu, Yongsheng and Lin Zhou (2007), "Rationalizability of choice functions by game trees." *Journal of Economic Theory*, 134, 548–556. [788]

---

Submitted 2011-3-23. Final version accepted 2012-10-1. Available online 2012-10-2.